Functional relevance of genetic variants in GPCRs

Dissertation zur Erlangung des akademischen Grades des
Doktors der Naturwissenschaften
(Dr. *rer. nat.*)

eingereicht im
Fachbereich Biologie, Chemie, Pharmazie der Freien Universität Berlin

vorgelegt von Master of Science

Annika Kreuchwig

aus Luckenwalde

2014

Freie Universität Berlin

FMP LEIBNIZ-INSTITUT FÜR MOLEKULARE PHARMAKOLOGIE

This work has been carried out from April 2010 – July 2014 in the group „Structural Bioinformatics and Protein Design“ under the supervision of Dr. Gerd Krause and Prof. Dr. Hartmut Oschkinat at the Leibniz-Institut für Molekulare Pharmakologie (FMP) in Berlin-Buch, Germany.

1^{st} reviewer: Dr. Gerd Krause

2^{nd} reviewer: Prof. Dr. Hartmut Oschkinat

Leibniz-Institut für Molekulare Pharmakologie (FMP)
Robert-Rössle-Straße 10, D-13125 Berlin

Thesis defence: February 23^{rd}, 2015

Acknowledgment

This thesis would not have been possible without the encouragement, cooperation and assistance of various people who deserve my acknowledgment. I would like to thank everyone who has helped me along the way. In particular:

I am exceptionally grateful to Dr. Gerd Krause, group leader of „Structural Bioinformatics and Protein Design“ at the Leibniz-Institut für Molekulare Pharmakologie (FMP) Berlin, for giving me the opportunity to work on this interesting topic, his guidance during my work and the pleasant working environment. I thank him for the systematic supervision, the great effort he put into training me in the scientific field and for giving me the chance for presenting my work at international conferences.

Further, I would like to thank Professor Dr. Hartmut Oschkinat for his support, the scientific discussions for collaborative projects and for being my university supervisor for this thesis.

I would like to express my sincere gratitude to Dr. Gunnar Kleinau, for the support and enthusiasm he has shown throughout the years, fruitful discussions and for initiating the project by building the first version of the database for sequence-structure-function-analysis of glycoprotein hormone receptor during his PhD thesis, which paved the way for my thesis.

Next, I would like to acknowledge Prof. Dr. Torsten Schöneberg who initiated the ADP receptor project and gave me the unique opportunity to cooperate on this very interesting topic. His friendly company, good ideas and deep interest have helped a lot to finish the project. I want to thank his group for their great effort performing *in vitro* mutagenesis and characterization of 1254 mutants, among them especially Maxi Cöster and Doreen Thor.

Additionally, I am grateful for having the opportunity to support several great interdisciplinary studies by sequence-structure-function-analyses. Great work and publications resulted from these contributions and therefore, I want to thank Prof. Dr. Dr. h.c. Stefan H.E. Kaufmann, Dr. Enno Klussmann and Dr. Ralf Schülein and their respective working groups.

During the last years many colleagues have become friends. I am thankful to my office mates, Dr. Gunnar Kleinau, Dr. Catherine L. Worth, Dr. Marcel Jurk, Jonas Protze, Dr. Anita Kinne, my sister Franziska, Katrin M. Hinz for support and thoughtful discussions in the office. In

addition to them, I acknowledge the whole group for the stimulating, always fun and warm, almost family-like atmosphere. My sincere thanks go to my friends Inna, Ann-Karin and Claudia for their encouragement whenever I need them, for scientific discussions as well as for gossip and activities after work.

I enjoyed my responsibilities as PhD representative and want to thank my fellows, Anja Voreck, Inna Hoyer and Karl Sydow for the good time and organization of many events, BBQ sessions, Autumn School and for enforcing PhD related rights. It was always fun working with you.

Finally, I want to thank all those who have contributed to finalize this thesis by reviewing parts or the whole text. Especially noteworthy are the following, Dr. Gerd Krause, Dr. Claudia Rutz and my sister Franziska.

I am lucky to have wonderful friends who have helped me through the last few years. I would also like to thank my family – my parents and my sister for all their support and advice and belief that I was capable of taking on the PhD. And most of all, I want to thank Lars, who has been endlessly patient, kind, loving and always there.

Graphical Abstract

Abstract

G protein-coupled receptors (GPCRs) are the key therapeutic targets for thirty to forty percent of all current marketed pharmaceutical drugs developed to cure many acute or chronic diseases. Investigations of naturally occurring in addition to mutations gathered from *in vitro* mutagenesis studies provide insight into the causes of human genetic diseases and provide novel perspectives for pharmacological strategies targeted at either mutant or wild type GPCRs. A GPCR subfamily, the glycoprotein hormone receptors (GPHRs) and the nucleotide receptor $P2Y_{12}$ were in the focus of this thesis. Naturally occurring pathogenic mutations in GPHRs have been identified as the cause of several endocrine diseases. One member of this subfamily, the thyrotropin receptor, is a key regulator of thyroid function, and the most prominent disorders related to this receptor are congenital hyper- or hypothyroidism and thyroid cancer. Activation of the $P2Y_{12}$ triggers platelet aggregation and thus this receptor plays a crucial role upon injury or thrombosis. Defects in the gene encoding for the $P2Y_{12}$ are often responsible for congenital bleeding disorders. Structural and functional studies of these receptors that elucidate the relevance of particular genetic variants pave the way for deriving more general, overarching mechanisms for other GPCRs. The structural elucidation of GPCRs facilitates the overall understanding of the functional and/or structural importance of certain residues. During recent years, family A GPCRs have been characterized in various studies thus generating a huge amount of data, which will even continue to increase in the future. This particularly concerns: i) the number of available functional data from mutagenesis studies, ii) the deposition and analysis of clinically relevant sequence variants, iii) the availability of complete genomes and thus sequence data of various species in the course of genome projects, and iv) the further structural elucidation of GPCRs by X-ray crystallographic analysis or nuclear magnetic resonance spectroscopy.

Tools for the analysis of these comprehensive and non-uniformly stored data are rare, however necessary to analyze the impact of genetic variants on diverse cellular processes. Affected are, for example, ligand binding, receptor expression, G protein coupling, receptor desensitization and receptor recycling. The combination of mutagenesis data with sequence and structural information allows for the identification of receptor modifications altering wild type receptor function. Furthermore, amino acid variations that are tolerated by the biological system, leaving the receptor function unchanged, can be determined. The computational approaches developed in this thesis for linking and unifying these completely different datasets create

extremely valuable resources facilitating the visualization, analysis and extraction for expert as well as non-expert users.

A mutation database for the GPHRs has been decisively advanced (SSFA-GPHR, available at http://www.ssfa-gphr.de). In addition, key changes aimed for improved user friendliness and the expansion to a comprehensive information platform have been implemented to increase its impact to the research community. In particular, implementation of novel tools, extension of the dataset and additional functionalities for data visualization and interpretation have proven to be extremely helpful in guiding and explaining experimental data, in order to understand the molecular reasons for impaired receptor function. Two application examples will be discussed, with both relating to the thyrotropin receptor, which has been the focus of extensive experimental studies over the last 25 years. In the first example, the database has proven to be essential for detecting the molecular basis for naturally occurring gain- or loss-of-function mutations by enabling the comparison of data from its homolog receptors, several species and different experimental approaches. Therefore, a tool for exploring and recognizing the spatial interrelationships between amino acid side chains in an interactive manner has been developed. The combination of functional characterization of known protein variants, statistical sequence analysis and structural evaluation, has shaped the next application. The study revealed pairs of interacting residues either stabilizing the basal or active receptor state of glycoprotein hormone receptors. A set of polar residues in transmembrane helices 2, 3, 6 and 7 are most likely involved in stabilizing the inactive state, while a mainly polar interface between transmembrane helices 5 and 6 is presumably important in holding the active state. Conservation of these features throughout family A GPCRs suggests their fundamental role in regulating receptor function.
Since no complete crystal structures have been solved for the GPHRs yet, homology models serve as a proxy and are constantly improved by novel structural and functional information. Additionally, fragments, such as the recently published crystal structure of the N-terminal region of the follitropin receptor in complex with its hormone follitropin, are taken into account. This structure, for instance, permitted homology modeling and subsequent systematic analysis of structural and functional data to reveal distinct residues, which are crucial for hormone recognition and binding at the N-terminal extracellular region of the homologous thyrotropin receptor. Allowing the user to further illuminate these contacts and interactions between the hormone and their respective receptors, an interactive interface analysis tool has been developed for the web application.
In general, homology models still play an important role in GPCR-related research, as crystal structures have been solved for only a small fraction of the huge family. The general fold is preserved, however what often hampers the template selection is the low sequence similarity. Therefore, a web accessible pipeline for GPCR homology modeling, the GPCR-Sequence-Structure-Feature-Extractor (SSFE, available at http://www.ssfa-7tmr.de/ssfe) that stores the template predictions, sequence alignments, identified sequence and structure motifs and homology models for 5025 family A GPCRs, has been developed. The specificity and unique feature of this method, in contrast to other available approaches, is that template selection is carried out for each transmembrane helix separately rather than for the receptor as a whole.

Template selection of the helix fragments is performed sequentially, and finally the helix bundle is assembled forming an overall model. Other published methods for GPCR homology modeling, employing one crystal structure template for the modeling process, do not benefit from the structural diversity of all previously published GPCR crystal structures. The method presented in this thesis is significantly faster in computing the homology models by providing similar or even more accurate results. Accuracy was determined by the calculation of the root mean square deviation of the transmembrane region of the homology models to its later published crystal structure.
The investigation of molecular effects caused by genetic variations can be facilitated by studying the evolutionary relationship of extant species. As a model protein a GPCR for Adenosine diphosphate (namely $P2Y_{12}$) was selected. Subsequently, the evolutionary conservation of 77 ortholog sequences (same protein but different species) was compared to the characterization of a comprehensive *in vitro* mutant library (site-saturation mutagenesis of every possible substitution at 66 contiguous positions, total of 1254 mutants). For the evaluation of both datasets, a web-based system ($P2Y_{12}$ mutant library; `http://www.ssfa-7tmr.de/p2y12`) was implemented to assess the correlation between the *in vivo* occurrence of variants in the ortholog dataset and the *in vitro* function of the characterized mutants. The calculation revealed a high correlation between *in vivo* and *in vitro* data (> 90%) which led to the conclusion that ortholog sequence data are sufficient to predict the functional relevance of individual positions and mutations not only within GPCRs, but likely for many conserved proteins as well. The significance of this approach will even increase in the future, since, due to various genome sequencing projects the amount of available sequence data, and thus the need for new technology to analyze this data will further rise.

GPCR function is shaped by a long evolutionary process characterized by mutation and natural selection. The likelihood that a particular residue or motif is responsible for proper GPCR function correlates to the degree of conservation amongst their ortholog sequences. The combination of evolutionary information, *in vitro* functional characterization, and structural evaluation of missense mutations, is an extremely powerful approach to interpret the correlation between genetic variations, their molecular causes, and altered receptor function. Finally, the combination of the concepts and tools described in this thesis are key steps towards the realization and implementation of a comprehensive platform for the analysis of sequence-structure-function relationships of all family A GPCRs. This ongoing project aims to provide sequences, structural and evolutionary information for a dataset comprising of as many as 20 500 family A GPCRs, facilitating i) the statistical analyses of evolutionary constraints at every helix position of a certain GPCR within this dataset, ii) the derivation of structural or functional reasons leading to variability or conservation of particular residues, iii) the tracing of conformational changes in-between activity states, and iv) binding pocket characterization and comparison.

Kurzfassung

G-Protein-gekoppelte Rezeptoren (GPCR) stellen bedeutende pharmakologische Zielstrukturen für die Wirkstoffforschung dar. Ihre Bedeutung zeigt sich vor allem darin, dass dreißig bis vierzig Prozent aller auf dem Markt befindlichen Arzneimittel direkt oder indirekt über diese Rezeptorfamilie wirken und zur Therapie der unterschiedlichsten Erkrankungen eingesetzt werden. Untersuchungen von natürlich vorkommenden sowie Mutationen aus *in vitro*-Mutagenesestudien sind daher unabdingbar und geben Einblicke in die Ursachen für humane genetische Erkrankungen. Weiterhin ermöglichen sie perspektivisch die Entwicklung pharmazeutischer Strategien an defekten sowie wildtypischen GPCR. Im Fokus dieser Arbeit stehen eine GPCR-Unterfamilie, die Glykoprotein-Hormon-Rezeptoren (GPHR) und der Nukleotidrezeptor $P2Y_{12}$. Eine Vielzahl an natürlich vorkommenden Mutationen konnte bereits als Ursache von endokrinen Erkrankungen identifiziert werden. Der Thyreotropin-Rezeptor, ein Mitglied der GPHRs hat Schlüsselfunktionen in der Regulation der Schilddrüsenfunktion. Die häufigsten Störungen dieses Rezeptors führen zu kongenitaler Hyper- oder Hypothyreose. Der $P2Y_{12}$ Rezeptor reguliert die Thrombozytenaggregation und spielt dadurch eine wichtige Rolle bei Verletzungen der Gefäßwand oder bei Thrombose. Angeborene Defekte in diesem Rezeptor führen häufig zu Blutgerinnungsstörungen. Spezifische Studien an einzelnen GPCRs verbessern das Verständnis von generellen Mechanismen und ermöglichen die Übertragung einzelner Merkmale auf andere Rezeptoren dieser Familie. Die strukturelle Aufklärung dieser Rezeptoren ermöglicht es, Erkenntnisse über die funktionelle und strukturgebende Relevanz einzelner Aminosäurepositionen des Rezeptors zu gewinnen. In den letzten Jahren wurden in diversen Studien zur Charakterisierung dieser Rezeptorfamilie riesige Datenmengen generiert, welche in Zukunft weiterhin rasant ansteigen werden. Dies betrifft vor allem, i) die Anzahl verfügbarer funktionaler Daten aus Mutagenesestudien ii) die Speicherung und Analyse von klinisch relevanten Sequenzvarianten, iii) die Verfügbarkeit vollständiger Genome und dadurch Sequenzdaten diverser Spezies im Zuge von Genomprojekten, und iv) die weitere Strukturaufklärung von GPCRs durch Kristallstrukturanalysen oder Kernspinresonanzspektroskopie.

Die funktionale Charakterisierung von Mutationsdaten und die Analyse natürlich vorkommender pathogener Veränderungen helfen bei der Aufklärung der Funktionsweise von GPCR. Werkzeuge zur Analyse dieser umfassenden und uneinheitlich gespeicherten Daten sind rar, jedoch nötig, um den Einfluss genetischer Varianten auf unterschiedliche zelluläre Prozesse zu analysieren. Beeinflusst werden beispielsweise, die Ligandenbindung, Rezeptorexpression,

G-Protein-Kopplung, Rezeptor-Desensitisierung, wie auch der Rücktransport der Rezeptoren zur Zelloberfläche. Die Kombination der Mutagenesedaten mit Sequenz- und Strukturinformationen ermöglicht es Rezeptormodifizierungen zu identifizieren, welche die natürliche Funktion einschränken. Weiterhin können auch Modifizierungen, die von dem System toleriert werden und das wildtypische Profil des Rezeptors nicht beeinflussen, ermittelt werden. Die in dieser Arbeit vorgestellten bioinformatischen Methoden zur Verknüpfung und Vereinheitlichung dieser grundverschiedenen Datensätze ermöglichen fachkundigen aber auch nicht fachkundigen Nutzern, die Visualisierung, Analyse und Extrahierung von Daten.

Eine Mutationsdatenbank für eine GPCR-Unterfamilie, den GPHR, konnte im Zuge dieser Arbeit entscheidend weiterentwickelt werden. Zudem stand die Nutzerfreundlichkeit und der Ausbau als umfangreiche Informationsplattform im Vordergrund, um einem größeren Anwenderkreis gerecht zu werden (SSFA-GPHR, verfügbar unter `http://www.ssfa-gphr.de`). Im Detail wurden neue Werkzeuge zur Analyse und Visualisierung programmiert sowie der Datenumfang ausschlaggebend durch die Integration von Mehrfachmutationen erweitert. Zusätzliche Funktionalitäten für die Datenvisualisierung und Interpretation zeigten sich als besonders hilfreich, sowohl für die Entwicklung neuer, als auch bereits durchgeführter Experimente. Mit Hilfe dieser Datenbank lassen sich die molekularen Ursachen für abnormale Rezeptorfunktionen aufdecken. Zwei Anwendungsstudien, die dies illustrieren, werden in dieser Arbeit diskutiert. Dabei beschränken sich beide Studien auf den Thyreotropin-Rezeptor, welcher in den letzten 25 Jahren und bis heute Gegenstand umfangreicher experimenteller Studien war und ist. In der ersten Studie wird deutlich, dass die Datenbank essentiell für die Erkennung molekularer Zusammenhänge natürlich vorkommender „Gain-“ und „Loss-of-function-“ Mutationen ist. Dies konnte durch den Vergleich von Daten homologer Rezeptoren unterschiedlicher Spezies sowie verschiedener experimenteller Ansätze in Kombination mit struktureller Information unterstützt werden. Technisch wurde dafür ein neues Werkzeug entwickelt, welches die Projektion von funktionalen Daten auf die dreidimensionale Struktur ermöglicht und dadurch die Analyse des räumlichen Zusammenspiels der Aminosäureseitenketten fördert. Das zweite Anwendungsbeispiel wird durch die Verknüpfung funktionaler Daten bekannter Proteinvarianten mit statistischer Sequenzanalyse und Strukturauswertung charakterisiert. Im Laufe der Studie konnten Interaktionspartner unter den Aminosäureresten identifiziert werden, die entweder den basalen oder den aktiven Rezeptorzustand von GPHR stabilisieren. Einerseits konnte ein polares Cluster von Aminosäuren zwischen den Transmembranhelices 2, 3, 6 und 7 ermittelt werden, welches sehr wahrscheinlich an der Stabilisierung des inaktiven Zustandes des Rezeptors beteiligt ist. Wohingegen hauptsächlich hydrophobe Interaktionen am Interface zwischen Transmembranhelix 5 und 6 für die Stabilisierung der aktiven Rezeptorkonformation zuständig sind. Die Konservierung dieser Interaktionen innerhalb der Familie A der GPCR weist auf ihre fundamentale Bedeutung für die Regulation der Rezeptorfunktion hin.
Da bisher noch keine vollständigen Kristallstrukturen der GPHR existieren, werden in der Datenbank Homologiemodelle verwendet, welche kontinuierlich durch neu bekanntwerdende Struktur- oder Funktionsdaten verbessert werden. Dabei werden auch Fragmente, wie die Struktur der Ektodomäne des Follitropin-Rezeptors in Komplex mit seinem Hormon Follitro-

pin verwendet. Dies ermöglicht die Homologiemodellierung und anschließende Analyse der Thyreotropin-Komplexbildung für den homologen Thyreotropin-Rezeptor. Dadurch konnten Aminosäurereste identifiziert werden, die ausschlaggebend für die Hormonerkennung sowie die Hormonbindung an die Ektodomäne dieses Rezeptors sind. Um die Interaktionsschnittstellen zwischen Hormon und Ektodomäne näher analysieren zu können, wurde ein interaktives Tool entwickelt, dass dem Nutzer die Möglichkeit gibt, Kontakte und Interaktionspartner zwischen dem Hormon und dem jeweiligen Rezeptor aufzudecken.
Homologiemodelle spielen in der GPCR-Forschung auch heute noch eine sehr große Rolle, denn bisher wurden Kristallstrukturen nur für einen sehr kleinen Bruchteil der großen Proteinfamilie gelöst. Aus diesem Grund wurde die frei zugängliche Plattform zur automatischen GPCR-Homologiemodellierung, „GPCR-Sequence-Structure-Feature-Extractor“ (SSFE, verfügbar unter `http://www.ssfa-7tmr.de/ssfe`), entwickelt. Diese generiert und speichert die Template-Vorhersagen, Sequenzalignments, die identifizierten Sequenz- und Strukturmotive sowie Homologiemodelle für 5025 Mitglieder der Familie A der GPCR. Die Besonderheit und das Alleinstellungsmerkmal dieser Methode im Gegensatz zu anderen verfügbaren Ansätzen ist, dass für jede Helix separat das jeweils beste Template gewählt wird. Die Selektion der Template-Strukturen wird nacheinander durchgeführt und anschließend werden die Helixfragmente zu einem Gesamtmodell zusammengefügt. Andere bekannte Methoden verwenden eine Kristallstruktur als Template für ein komplettes Modell. Dadurch können sie nicht von der strukturellen Vielfalt aller bereits publizierten GPCR Kristallstrukturen profitieren. Homologiemodelle, die mit der hier vorgestellten Methode generiert wurden, zeigten sich im Vergleich zu anderen Vorhersagemethoden, schneller in der Berechnung der Modelle und lieferten dabei ähnlich akkurate und in einigen Fällen sogar präzisere Resultate. Die Genauigkeit konnte anhand der mittleren quadratischen Abweichung des Transmembranbereiches der Homologiemodelle zur jeweils später publizierten Kristallstruktur überprüft werden.
Die funktionelle Relevanz genetischer Varianten (Mutationen) kann mit Hilfe von Analysen der evolutionären Zusammenhänge bestehender Spezies aufgedeckt werden. Um diese Hypothese zu überprüfen, wurde ein Modellprotein aus der Familie A der GPCR, der Adenosindiphosphat-Rezeptor $P2Y_{12}$, gewählt. Im folgenden Schritt konnte die evolutionäre Konservierung von 77 Orthologen (gleiches Protein unterschiedlicher Spezies) mit einem Datensatz von *in vitro* funktional charakterisierten Mutanten verglichen werden. Es wurde eine sättigende Mutagenese durchgeführt, das heißt an jeder Position wurden alle anderen 19 Aminosäurevarianten eingefügt und anschließend funktionell charakterisiert. Insgesamt entstanden so 1254 Mutationen. Für den Vergleich beider Datensätze und die anschließende statistische Analyse wurde eine web-basierte Plattform implementiert („$P2Y_{12}$ Mutant Library“; `http://www.ssfa-7tmr.de/p2y12`). Mit Hilfe dieser Plattform lassen sich Übereinstimmungen zwischen der evolutionären Konservierung bzw. Variabilität und den *in vitro*-Mutagenesedaten finden. Die Studie zeigte, dass anhand der Orthologdaten die Funktion von mehr als 90% der Mutanten korrekt vorhergesagt werden konnte. Schlussfolgernd können demzufolge orthologe Sequenzdaten verwendet werden, um die funktionelle Relevanz einzelner Positionen und Mutationen sowohl für GPCR als auch für andere konservierte Proteine vorherzusagen. Dieser Ansatz wird zukünftig noch wichtiger, da im Zuge

verschiedener Genomsequenzierungsprojekte die Anzahl der verfügbaren Sequenzdaten steigt und dadurch auch die Nachfrage nach neuen Technologien zur Analyse dieser Daten.

Die Funktionsweise von GPCR wurde im Laufe der Evolution durch viele Prozesse, wie zum Beispiel Mutationsereignisse und natürliche Selektion, geprägt. Die Wahrscheinlichkeit, dass ein bestimmter Aminosäurerest oder ein Motiv wichtig für die Aufrechterhaltung der GPCR-Funktionen ist, korreliert mit dem Grad der Konservierung innerhalb der orthologen Sequenzen. Die Kombination von evolutionärer Information, funktioneller Charakterisierung und struktureller Auswertung von „missense"–Mutationen ist ein sehr vielversprechender Ansatz, um die Korrelation zwischen genetischer Variante, ihrer molekularen Zusammenhänge und der veränderten Rezeptorfunktion zu interpretieren. Zusammengefasst, stellen die in dieser Arbeit beschriebenen Konzepte und Werkzeuge einen entscheidenden Schritt zur Entwicklung einer allumfassenden Plattform zur Analyse von Sequenz-Struktur-Funktionsbeziehungen von GPCR der Familie A dar.
Ein Ziel über diese Arbeit hinaus ist es, eine Ressource zu bieten, die Sequenzen, Strukturen und evolutionäre Zusammenhänge für GPCR der Familie A aus unterschiedlichen Spezies (insgesamt umfasst dieser Datensatz 20 500 Sequenzen) zur Verfügung stellt und gleichzeitig tiefergehende Analysen zulässt. Diese web-basierte Plattform wird momentan entwickelt und ermöglicht künftig i) die statistische und evolutionäre Auswertung an jeder einzelnen Helixposition eines bestimmten GPCR, ii) die Ableitung struktureller oder funktioneller Determinanten, die zur Variabilität bzw. Konservierung einzelner Positionen führen, iii) das Verfolgen konformationeller Änderungen zwischen verschiedenen Aktivitätszuständen der Rezeptoren und iv) die Charakterisierung und der Vergleich von Bindungstaschen in unterschiedlichen GPCR.

List of publications

The majority of this thesis has been published in peer reviewed journals. Please find the articles corresponding to the parts of this thesis below.

1. Annika Kreuchwig, Gunnar Kleinau, Franziska Kreuchwig, Catherine L Worth, Gerd Krause
   Research resource: Update and extension of a glycoprotein hormone receptors web application. *Molecular Endocrinology*, Apr 2011, 25, 707-12.
   [Section 2.1, page 10]
2. Gunnar Kleinau, Annika Kreuchwig, Catherine L. Worth, Gerd Krause
   An interactive web-tool for molecular analyses links naturally occurring mutation data with three-dimensional structures of the rhodopsin-like glycoprotein hormone receptors. *Human Mutation*, Jun 2010, 31, E1519-25.
   [Section 2.2, page 19]
3. Gunnar Kleinau, Inna Hoyer, Annika Kreuchwig, Ann-Karin Haas, Claudia Rutz, Jens Furkert, Catherine L. Worth, Gerd Krause, Ralf Schülein
   From molecular details of the interplay between transmembrane helices of the thyrotropin receptor to general aspects of signal transduction in family A G-protein-coupled receptors (GPCRs). *Journal of Biological Chemistry*, Jul 2011, 286, 25859-71.
   [Section 2.3, page 28]
4. Catherine L. Worth, Annika Kreuchwig, Gunnar Kleinau, Gerd Krause
   GPCR-SSFE: a comprehensive database of G-protein-coupled receptor template predictions and homology models. *BMC Bioinformatics*, May 2011, 23, 12:185.
   [Section 2.4, page 43]
5. Gerd Krause, Annika Kreuchwig, Gunnar Kleinau
   Extended and structurally supported insights into extracellular hormone binding, signal transduction and organization of the thyrotropin receptor. *PLoS One*, Dec 2012, 7, e52920.
   [Section 2.5, page 55]
6. Annika Kreuchwig, Gunnar Kleinau, and Gerd Krause
   Research resource: novel structural insights bridge gaps in glycoprotein hormone receptor analyses. *Molecular Endocrinology*, Aug 2013, 27(8):1357-63.
   [Section 2.6, page 68]

7. Maxi Cöster, Doreen Wittkopf, Annika Kreuchwig, Gunnar Kleinau, Doreen Thor, Gerd Krause, Torsten Schöneberg
   Using ortholog sequence data to predict the functional relevance of mutations in G-protein-coupled receptors. *FASEB Journal*, Aug 2012, 26, 3273-81.
   [Section 3.1, page 77]

Contents

List of Figures

List of Abbreviations

Acronyms

2D	Two-dimensional
3D	Three-dimensional
Å	angstrom
ADP	Adenosine diphosphate
AhR	Aryl hydrocarbon receptor
AKAP	A-kinase anchor protein
AMBER	Assisted Model Building with Energy Refinement
Ba – We	Ballesteros Weinstein
CAM	Constitutively activating mutation
CASP	Critical Assessment of Techniques for Protein Structure Prediction
CG	Chorionic gonadotropin
CXCR4	C-X-C chemokine receptor type 4
D/D	Dimerization and Docking
ECL	Extracellular loop
FSH	Follicle-stimulating hormone, Follitropin
FSHR	Follicle-stimulating hormone receptor, Follitropin receptor
GDP	Guanosine diphosphate
GPCR	G protein-coupled receptor

GPCR – SSFE	GPCR-Sequence-Structure-Feature-Extractor
GPCRDB	G protein-coupled receptor database
GPCRRD	GPCR Research Database
GPHR	Glycoprotein hormone receptor
GRK	G protein-coupled receptor kinase
GTP	Guanosine-5'-triphosphate
G protein	Guanosine nucleotide-binding protein
I – TASSER	Iterative Threading Assembly Refinement
ICL	Intracellular loop
IUPHAR	International Union of Basic and Clinical Pharmacology
LH	Luteinizing hormone, Lutropin
LHCGR	Luteinizing hormone/Chorionic gonadotropin receptor
LRR	Leucine-rich-repeat
LRRD	Leucine-rich-repeat domain
MeS – ADP	2-(methylthio)adenosine 5'-diphosphate
MSA	Multiple Sequence Alignment
N – ECR	N-terminal extracellular region
NaVa	GPCR Natural Variant database
NMR	Nuclear magnetic resonance
PAS	Per-ARNT-Sim
PDB	Protein Data Bank
*PHP*5	PHP Hypertext Preprocessor scripts version 5
PKA	Protein kinase A
RGS	Regulators of G protein signaling
RMSD	root-mean-square deviation
SSFA – GPHR	Sequence-Structure-Function-Analysis of Glycoprotein Hormone Receptors
TASSER	Threading Assembly Refinement
TCDD	2,3,7,8-Tetrachlorodibenzo-p-dioxin

TMH	Transmembrane helix
TSH	Thyroid-stimulating hormone, Thyrotropin
TSHR	Thyroid-stimulating hormone receptor, Thyrotropin receptor
URL	Uniform Resource Locator
wt	wild type
WWW	World Wide Web

Amino Acids

Ala, A	Alanine
Arg, R	Arginine
Asn, N	Asparagine
Asp, D	Aspartic acid
Cys, C	Cysteine
Gln, Q	Glutamine
Glu, E	Glutamic acid
Gly, G	Glycine
His, H	Histidine
Ile, I	Isoleucine
Leu, L	Leucine
Lys, K	Lysine
Met, M	Methionine
Phe, F	Phenylalanine
Pro, P	Proline
Ser, S	Serine
Thr, T	Threonine
Trp, W	Tryptophan
Tyr, Y	Tyrosine
Val, V	Valine

CHAPTER 1

General introduction

This chapter presents the motivation behind this thesis, its objectives and the strategic approach to the problem presented. The chapter closes with a structural overview of the report.

1.1 Historical aspects - G protein-coupled receptors and their relationship to the Nobel Prize

Stimuli such as light, odors or flavors are recognized by sensors in our eyes, nose and mouth and illicit a response in our body. Similar sensors exist within our body which detect hormones and neurotransmitters. But who is responsible for sensing those signals and how is extracellular information converted into intracellular chemical messages between cells in our body? Research related to these and similar questions have primarily been performed over the past 40 years and have a rather long relationship with the Nobel Prize. G protein-coupled receptors (GPCR) are receptive molecules located in the plasma membrane that regulate almost all physiological processes from vision, taste, sense of smell, muscle movement, metabolism and immune regulation to reproduction. After receiving a stimuli, their task is it to convey a signal into the cell and to switch on some yet turn off other reactions in order to finally deliver their message to receive an appropriate cellular and physiological response. Today, it is known that there are more than 1000 human genes which encode the GPCR superfamily [84, 87]. However, the process was rather long which resulted in the confirmation that the proteins which pick up a wide variety of signals actually belong to the same class, as they use the same basic mechanism for signaling via heterotrimeric guanosine nucleotide-binding proteins (G proteins) inside the cell. The story of GPCRs started in the 1876 when rhodopsin was identified as a light-sensitive pigment by Franz Christian Boll [8]. In 1933, its ligand retinal was reported [88]. At the beginning of the 20th century Dale [15] and Langley [44] shaped the history of GPCRs by introducing the concept of ligand-activated receptors. This statement was initially not widely accepted, until 1968 when Robert J. Lefkowitz was one of the first to demonstrate the existence of receptors

seated in the plasma membrane of cells by use of radioactively labeled ligands in receptor binding studies [47, 48]. Simultaneously, the principle of signal transduction in the interior of cells by second messengers was first described by Earl W. Sutherland in 1967 [67] and honored with the Nobel Prize in Physiology or Medicine in 1971. In the years to follow, this resulted in the description of the ternary complex model for the mechanism of receptor activation by an extracellular ligand and transmission via the transmembrane region, thus inducing the coupling of intracellular G proteins to form an activated signaling unit [17]. "For their discovery of G-proteins and the role of these proteins in signal transduction in cells" Alfred G. Gilman and Martin Rodbell were awarded the Nobel Prize in Physiology or Medicine in 1994. In 2004, the Nobel Prize was awarded to Richard Axel and Linda B. Buck for their discovery of a specialized class of GPCRs that control the olfactory system. In the following years, Robert J. Lefkowitz and co-workers research focused on the β-adrenergic receptor. Following the first hurdle of solubilizing the receptor from the membrane [9], Brian Kobilka managed to clone the gene and derived its genetic sequence [4]. Seven alpha helices, as secondary structural elements and its sequence homology to the well-known rhodopsin, led the researchers to the assumption of the existence of an entire class of proteins with a close structural relationship responding to a wide range of signals [54]. The Nobel prize in 2012 acknowledged the general importance of GPCRs and particularly emphasized the contributions of Lefkowitz and Kobilka to the elucidation and characterization of GPCRs [64, 69].

1.2 Classification of GPCRs

In 2003, Fredriksson and colleagues [22] established a classification system for GPCRs based on sequence analysis and phylogenetic investigation of the human genome. The analysis revealed that the human receptors cluster into five main families that are termed rhodopsin-like, secretin, glutamate, adhesion and frizzled/taste2. Today, the most-commonly used classification system has been established in the GPCRDB [87], dividing the GPCRs in six families. Approximately 80% of all human GPCRs are grouped into the rhodopsin-like family (family A). In addition, the GPCRs are classified as secretin-like (family B), metabotropic glutamate receptors (family C), pheromone receptors (family D), cAMP receptors (family E) and the much smaller frizzled/smoothened family (family F). Most GPCR families are found in mammalian species (A, B, C and F), while family D (found only in fungi) and family E (found only in *Dictyostelium discoideum*) are species-specific [16]. GPCR analyses described in the following sections concentrate on family A GPCRs with a special focus on glycoprotein hormone receptors (GPHRs) and P2Y nucleotide receptors.

1.3 GPCR topology

The general topology of family A GPCRs consists of seven transmembrane helices connected by three extracellular loops (ECLs) and three intracellular loops (ICLs). The structure of the transmembrane bundle is highly conserved; the loops vary in size and in the composition of secondary structure elements for some of the loops [85]. In addition, the size of the

extracellular N terminus can range between relatively short to large N-terminal regions, e.g. GPHRs, possess a large ectodomain. On the intracellular side, just before the C terminus, most family A GPCRs have an eighth α-helix located parallel to the membrane. Depending on the receptor, the ligand binds either extracellularly or within the transmembrane helix bundle. The intracellular region is involved in binding downstream interaction partners, including G proteins and β-arrestin as the most prominent ones.

1.4 Structural elucidation of GPCRs

In 1993, Schertler and Hendersen presented the first projection map of rhodopsin via cryo-electron microscopy [74] demonstrating the configuration of the polypeptide chain which traverses the membrane seven times. In 2000, Palczewski solved the first crystal structure of rhodopsin at 2.8 angstrom ($\AA$) resolution in the inactive conformation with bound inverse agonist 11-*cis*-retinal [60, Protein Data Bank (PDB) code: 1F88]. The seven-transmembrane bundle, which sits in the membrane, is dynamic and assumes different conformations during the activation process. The receptor generally adopts its active conformation upon agonist binding leading to a shape change which results in an increased affinity for the G protein. The first partially activated GPCR structures were published in 2008 and include ligand-free opsin [63, PDB code: 3CAP] and opsin with bound transducin fragment [73, PDB code: 3DQB]. In 2011, Kobilka and co-workers published the first agonist-bound active-state structure of β_2-adrenergic receptor stabilized, by a nanobody imitating G protein-like behavior [65, PDB code: 3P0G]. Several months later, the whole transmembrane signaling complex of activated β_2-adrenergic receptor coupled to the heterotrimeric G protein was solved [65, PDB code: 3SN6]. Comparison of both activity states for several GPCRs revealed similar changes upon activation. Most changes occur on the intracellular side and include an outward movement of transmembrane helix (TMH) 6 and an α-helical extension of TMH5. The publication of GPCR structures and thereby the elucidation of mechanisms such as ligand recognition, binding, activation, conformational changes in the transmembrane region and G protein coupling, guides pharmacological studies in identifying targeted medications with fewer side effects. Until 2010, there were only a handful of GPCR crystal structures available, with an increasing fluency in the identification of new structures appearing ever since. Today, a repertoire of 19 different family A (rhodopsin-like) structures are accessible, with several existing in different activity states. Two structures have been recently solved for family B (secretin-like) and one for family F (frizzled/smoothened family) (see Figure 1.1).

1.5 GPCRs as drug targets

The therapeutic relevance of GPCRs relies on their functional mechanism and physical location in the plasma membrane which enables the transduction of extracellular signals into intracellular responses. Thirty to forty percent of all current marketed pharmaceutical drugs target GPCRs [21, 59] making them the key therapeutic target for many acute or chronic diseases, including cancer [45], cardiac dysfunction [19], asthma [11], diabetes [90], central nervous system diseases [13], gastrointestinal diseases [55], obesity [24], inflammation

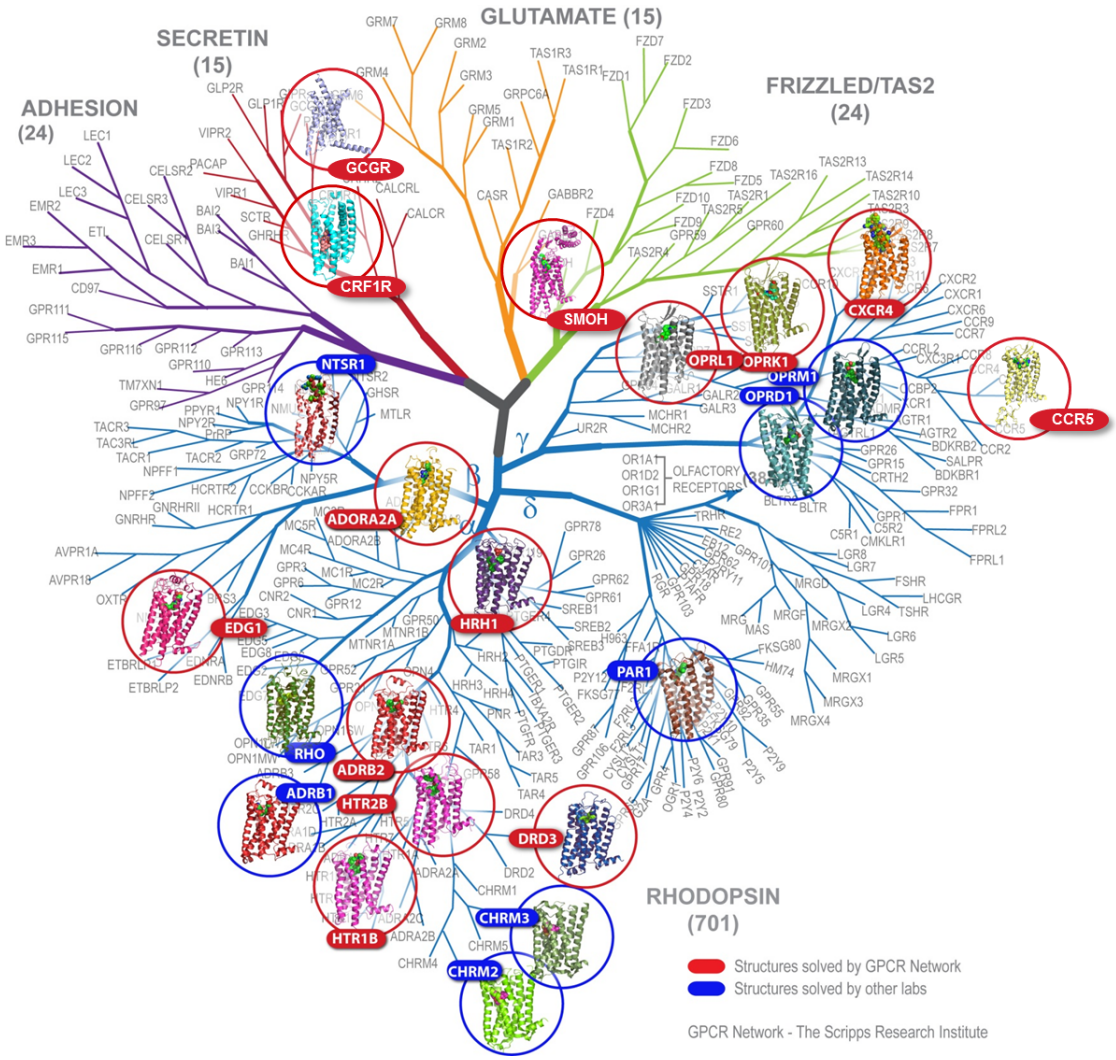

FIGURE 1.1 – GPCR family tree.

Based on phylogeny, the GPCRs can be grouped into five distinct families. Here, all currently known crystal structures are highlighted. Most structural information concerning rhodopsin-like GPCRs is known, however the first structures of other sub-families such as secretin and the frizzled family have only been recently published. The figure has been modified from the GPCR Network - The Scripps Research Institute.

[18, 49], and pain [1, 62]. One example is beta blockers which were discovered by the pharmacologist Sir James W. Black, who received the Nobel Prize for designing the first drugs blocking the β-adrenergic receptor. A primary part of this thesis focused on the nucleotide-like receptor $P2Y_{12}$, which is the target for the world's second-best selling drug clopidogrel which is used in anti-platelet therapy [28].

Recently, a new crystal structure has been identified to demonstrate the allosteric inhibition of chemokine signaling in the chemokine receptor CCR5 [79] by a small molecule termed Maraviroc. This, and other studies, will provide valuable implications in enabling structure-based drug discovery for the treatment of HIV-1 infection.

1.6 GPCR SIGNALING

Agonist binding at the extracellular side of GPCRs induces slight conformational changes and forms an activated receptor state which permits an association with its corresponding heterotrimeric G protein on the intracellular side. Within the cell, GPCR activation initiates a cascade of reactions (see Figure 1.2). The main signal transduction pathway of GPCRs is the G protein-dependent signaling, however, there is also evidence for G protein-independent signaling via β-arrestins [23, 61] or G protein-coupled receptor kinases [50]. For the classical pathway, the heterotrimeric G protein binds to the activated GPCR which triggers the GDP/GTP exchange within the Gα subunit. This leads to the dissociation of Gα subunit from G$\beta\gamma$ dimer permitting both to act upon various downstream effectors thus triggering unique signaling responses. The type of G protein activated by a particular agonist determines the further signaling route taken (functional selectivity). Activation of most GPCRs switch on more than one G protein and thereby initiates several signal transduction cascades, however there are several agonists known for their selective activation, e.g. cell type specific activation (reviewed in [51]). Inactivation of the receptor G protein complex and reassociation of the heterotrimeric G protein is enabled by the intrinsic GTPase activity of the Gα subunit, which converts GTP to GDP by hydrolyzation. The regulators of G protein signaling (RGS) proteins bind to Gα and are able to accelerate the intrinsic hydrolysis process [27]. Aside from the G protein dependent signal termination, signal attenuation can be driven by desensitization of the GPCR itself. This process is initiated by phosphorylation of the activated GPCR by G protein-coupled receptor kinases (GRKs) which leads to an increased affinity for β-arrestin binding. Binding of β-arrestin to a GPCR prevents G protein coupling by steric hindrances, thereby preventing further activation and initiates receptor internalization via clathrin-mediated endocytosis (reviewed in Wolfe & Trejo [91]).

1.7 GENETIC VARIANTS AND GPCRS IN DISEASE

Extensive analysis on the role of specific residues involved in receptor activation and/or ligand interaction provides insights on modifications which alter receptor function and parts of the receptor tolerating variations. Several studies *in vitro* [7, 25] or *in vivo* [2, 66] have tried to shed light on GPCR function and regulation. Often *in vitro* site-directed mutagenesis by the site-specific introduction of point mutations is used to analyze the importance of a particular residue for maintenance of the active or inactive receptor conformation, ligand binding or intracellular effector binding. Other *in vitro* methods include alanine or cysteine scanning and random mutagenesis. Each method has their advantages and disadvantages, but they commonly share several features such as that they examine only a small subset of functional properties and are not able to estimate the residue characteristics necessary for proper receptor functionality. For example, alanine scanning is not capable of predicting the functional effect of all other possible amino acid substitutions. With the example of the thyrotropin receptor (TSHR), alanine scanning demonstrated wild type function for the two residues Asn650Ala and Val656Ala in the extracellular loop 3 [12]. Subsequently, pathogenic mutations at these positions have been identified for Asn650Tyr, and Val656Phe, with both of

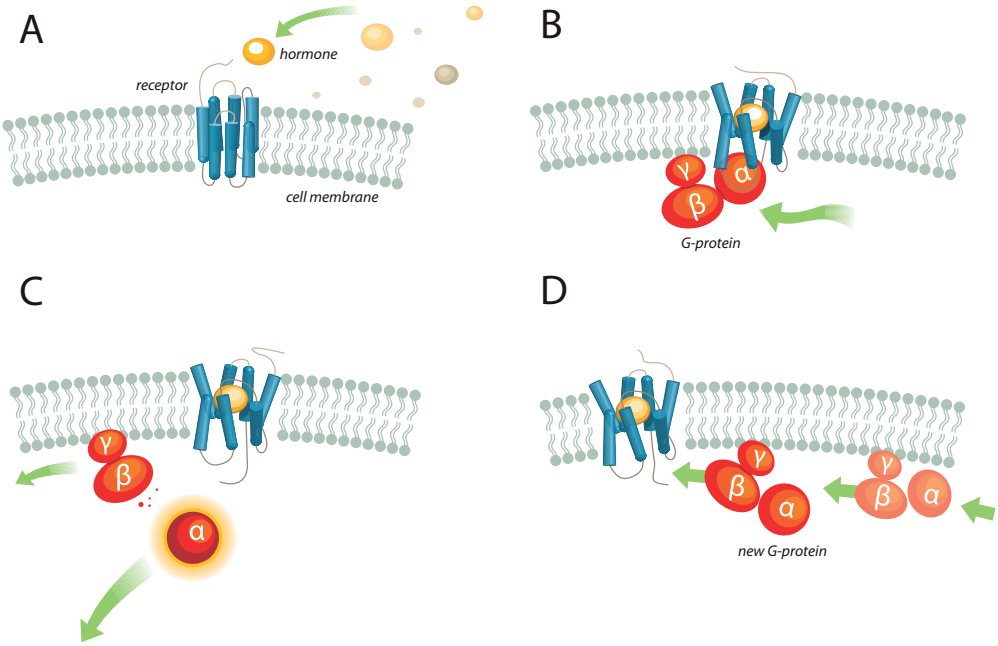

FIGURE 1.2 – GPCR signaling.

Upon extracellular ligand binding to a GPCR on the cell surface (A), a cascade of reactions is triggered inside the cell (B). This includes GDP/GTP exchange resulting in G protein subunit dissociation (α from $\beta\gamma$) (C). The activated receptor is able to bind further G proteins until the ligand detaches. Figure adopted from `http://www.nobelprize.org`, © The Royal Swedish Academy of Sciences [81].

them resulting in the constitutive activation of the TSHR [57, `http://www.ssfa-gphr.de`]. Interestingly, Small and colleagues [77] determined a substantial genetic variability in the GPCR superfamily ranging from single nucleotide polymorphisms (SNPs) to copy number variants. These variants differ in their impact on DNA level, protein function and finally in their effects on human phenotypes varying from neutral to disease-associated [34]. Naturally occurring variants responsible for monogenic diseases have been reviewed, amongst others, by Schöneberg et al. [75] and Tao [80]. In addition to those rare, disease-causing mutations, many natural variants have either no clinical or an unknown effect on protein function. It has been discussed as to whether the high degree of variability reflected through polymorphisms has stronger effects on adverse drug and ligand effects than on disease state [31]. For pathogenic mutations, one differentiates between gain-of-function mutations usually leading to constitutive receptor activation or enhanced ligand binding and loss-of-function mutations resulting in ligand resistance or reduced binding. For the receptors studied in this work, many activating or inactivating pathogenic mutations have been reported. For the GPHRs and particularly the TSHR, disruption or dysregulation of the receptor function eventually leads to thyroid disease including hypothyroidism, hyperthyroidism and thyroid cancer. The $P2Y_{12}$ receptor belonging to the purinergic receptors is involved in platelet aggregation. Receptor defects usually result in decreased aggregation responses upon activation via its ligand ADP or increased bleeding times upon injury [58].

1.8 DATABASE AND WEB SERVICES FOR GPCRS

There are a number of web resources available focusing on structure-function relationships of GPCRs. The GPCR database (GPCRDB), first published in 1998 [30], has continuously developed into an information system for GPCRs storing data on sequences, mutagenesis studies and ligand binding. During the course of the last 15 years, various tools such as an alignment builder and a tool to predict the effect of mutations *in silico* have been added [29, 87]. The mutagenesis studies are extracted from the literature either automatically or manually, and stored in the MuteXt repository or in the tinyGRAP database, respectively. The International Union of Basic and Clinical Pharmacology (IUPHAR) database for GPCRs collates and provides a comprehensive collection of peer-reviewed pharmacological, functional and pathophysiological information on human and rodent non-sensory GPCRs gathered from other resources which are linked at the appropriate places [26, 76]. Human-gpDB, another information resource, provides evidence for interactions between human GPCRs, G proteins and their downstream effectors [71]. Oligomer/dimer formation of GPCRs is curated manually and presented by the GPCR-OKB web resource [35].
The GPCR Natural Variants database (NaVa) integrates data on natural variants in human GPCRs. The database offers detailed information extracted from online databases, literature, and patents pertaining to each variant concerning location (DNA/amino acid), the involved nucleotides/amino acid, the average frequency of each allele, reported disease associations, and references.
Focused subfamily-specific databases and web services were developed for glycoprotein hormone receptors by SSFA-GPHR [36] and GRIS [82] providing homology models, sequences, alignments along with functional data extracted from the literature.
The GPCR Research database (GPCRRD) developed by Zhang and co-workers was the first web-service collecting GPCR-related spatial restraints derived from experimental data (mutagenesis studies, disulphide mapping distances, electron cryo-microscopy) to guide GPCR structure and function modeling [95]. Several years earlier, the same group applied their threading assembly refinement (TASSER) method to generate homology models of 907 human GPCRs [97]. GPCR-I-TASSER has been developed to combine the structural restraints stored in GPCRRD with the TASSER approach to derive high resolution molecular models for GPCRs (*unpublished data*).

1.9 EVOLUTIONARY CONSTRAINTS RESPONSIBLE FOR GPCR FUNCTION

The growing number of fully sequenced genomes of extant species provides us with a vast amount of GPCR sequences which can be used to identify either highly conserved sequence patterns responsible for GPCR functionalities or signatures which have been the subject of past and ongoing selection. The genetic repertoire of a particular receptor can be derived from the evolutionary history summarized in a multiple sequence alignment of ortholog species. GPCR function is shaped by a long evolutionary process characterized by mutation and natural selection. The continuous accumulation of non-deleterious mutations along with highly conserved sequence motifs reveals the functional relevance of distinct

residues in GPCRs. For detection, bioinformatic methods can be used for quantitative sequence comparisons. These are not only interesting between ortholog species, but also by combining different subfamilies or families to elucidate differences and commonalities for ligand binding, receptor activation and intracellular signaling. These kind of analyses facilitate the interpretation of disease-causing mutations in GPCRs. The likelihood that a particular residue or motif is responsible for proper GPCR function correlates to the degree of conservation among their ortholog sequences. The combination of evolutionary information, *in vitro* functional characterization and structural evaluation of missense mutations is an extremely powerful approach in interpreting the correlation between genetic variation, their molecular causes and altered receptor function [68, 78].

1.10 Aim and outline of this thesis

The aim of the work presented in this thesis was to gain insights into the structure-function relationships of family A GPCRs. The main focus was the deduction of structural and functional determinants underlying mechanisms such as ligand binding, molecular activation and G protein coupling. Of central importance in elucidating structure-function properties of the wild type receptor is the fact that amino acid side chain substitutions often modify the receptor phenotype. To understand the relationship between amino acid modification, structural change and molecular relevance, *in vitro* experimental and computational studies were combined to evaluate the molecular effects of genetic variations, providing insights into the causes of human diseases ultimately leading to novel pharmacological strategies targeted at either mutant and wild type GPCRs. Linking the amino acid sequence together with functional data and the structural characterization of GPCRs, will facilitate the overall understanding of the functional and/or structural importance of certain residues. The combination of platforms gathering experimental and pathogenic mutation data with computational approaches will be an extremely valuable resource for the analysis of the influence of genetic variants on ligand binding, receptor expression, G protein-coupling, receptor desensitization and receptor recycling as well as their effects on recognition and response to therapeutic agents. The published manuscripts discussed in this thesis can be divided by content into two chapters, namely Chapters 2 and 3.

Chapter 2 introduces the web application for Sequence-Structure-Function Analysis of glycoprotein hormone receptors (SSFA-GPHR) and elucidates its progressive and innovative development during the span of this thesis. Aside from novel tools and additional functionalities, two application examples, where the database has been proven to be essential in guiding experimental studies, are outlined. Since homology models play a significant role as a proxy for those receptors who do not have crystal structures, a web server for automatic template predictions and homology model building of GPCRs based on a multi-template and fragment based approach has been developed. The pool of crystal structures available in 2009 exhibited only a small subset of all possible structural variations of family A GPCRs. Systematic sequence and structural analysis revealed that some GPCRs contain sequence motifs causing particular structural features which are spread over different template crystal structures or which are not present in the templates which have been published to date. In

2011, this was the first method which integrated all structural information produced until then, by assembling the homology model based on the best matching template per helix.
Chapter 3 focuses on a site-saturation mutagenesis approach to prove that comparative sequence data are sufficient to predict the functional relevance of individual positions and mutations in GPCRs. The main part describes the methodological approach as a systematic proof of principle linking structural data, evolutionary sequence information and experimental findings of 1254 mutants covering 66 contiguous receptor positions. Furthermore, the section aims to demonstrate the developed tools and database to analyze the sequence-structure-function relationship of the ADP receptor to support novel ideas in the areas of pharmacological intervention, signal transduction and regulation.
Chapter 4 brings these discussions together, comparing the analysis of *in vitro* experimental studies with structural information as well as reflecting on the relevance of evolutionary sequence information for drug development and disease. As an excursus, collaborative projects from a different research field demonstrate how the methods for sequence-structure-function-analyses can be adapted to support studies on ligand-target interactions for other proteins of interest. Finally, ongoing work and future perspectives that emerge from the results of this thesis will be presented.

CHAPTER 2

Linking functional data with structural information of GPHRs

This chapter introduces the approach and continuous development of a mutation database for a GPCR subfamily, the glycoprotein hormone receptors (GPHRs), to deduce structural and functional determinants of mechanisms such as ligand binding and molecular activation up to G protein coupling. Besides implementation of novel tools for data visualization and interpretation [42, 43], the focus will be to demonstrate its utility for several applications elucidating GPCR-related mechanisms [37, 39]. Since no crystal structures have been solved for these GPCRs yet, homology models serve as a proxy and are constantly improved by novel structural and functional information [40]. Furthermore, a web accessible pipeline for GPCR homology modeling, the GPCR-Sequence-Structure-Feature-Extractor (SSFE, available at `http://www.ssfa-7tmr.de/ssfe`*) that stores the template predictions, sequence alignments, identified sequence and structure motifs and homology models for 5025 family A GPCRs [93], will be briefly explained.*

2.1 Sequence-structure-function-analysis of glycoprotein hormone receptors web application

2.1.1 *Synopsis*

The web application presented in the following publication aims to collate and link the functional characterization of receptor variants derived from experimental studies with structural information from homology models or crystal structures. The database is not only focused on the comprehensive collection, but also on the comparison and analysis of data from mutagenesis studies as well as naturally occurring mutations for the glycoprotein hormone receptors (GPHRs), a subgroup of rhodopsin-like GPCRs. The family includes

the thyroid-stimulating hormone receptor (TSHR), follicle-stimulating hormone receptor (FSHR) and luteinizing hormone/chorionic gonadotropin receptor (LHCGR), each binding their corresponding glycoprotein hormone: thyrotropin (TSH), follitropin (FSH), lutropin (LH) and chorionic gonadotropin (CG), respectively. Defects in these receptors, either via gain-of-function or loss-of-function mutations, are known to cause severe diseases such as hyper- and hypothyroidism and thyroid carcinoma for the TSHR.
A first version of this database and information source for GPHRs was published in 2007 by Kleinau and colleagues [36]. The basic idea of the platform called Sequence-Structure-Function-Analysis of GPHRs (SSFA-GPHR) is to offer a comparison and investigation of these homologous receptors in order to deduce the causes of associated diseases. The functional information extracted from the literature covers mutations which occur naturally as well as derived from *in vitro* mutagenesis studies. The unique feature of this database is the comparison of functional effects of particular mutations and linking these to structural properties. To enable these comparative analyses as well as classification of data from GPHR subtypes and different experimental approaches, the functional data are converted into unified scaled percentage values (wild type 100%).
The aim of this work was the redesign and development of an improved version with extended functionalities. Therefore, new tools for data visualization and interpretation were implemented. Novel features in the second database release include:

- interactive 3D-structure-based search for functional data and mutant information,
- extension of the dataset, aside from single mutations, implementation of double and triple mutations,
- structural morphings between basal and activated receptor conformations allowing for the tracing of changes in amino acid interactions during activation,
- search options directly in the 3D structure or via 2D – Snake-like – Plots,
- improved handling and features for data-analyses.

The goal was to establish a database, which would be flexible and could accommodate heterogeneous data from various mutagenesis studies published in the literature. The application SSFA-GPHR is built by combining an Apache web server (`http://www.apache.org`) hosting PHP Hypertext Preprocessor scripts (PHP 5) and a relational database management system (MySQL 5). Three-dimensional structures are displayed using the JavaScript-based structure viewer Jmol [33]. For projecting functional information of certain amino acid positions onto three-dimensional structure, homology models have been generated. Since no full receptor structure is yet available for GPHRs, published fragments (PDB codes: 1XWD, 3G04) of the leucine-rich-repeat (LRR) domain and homologous templates for the serpentine domain (in inactivating (PDB code: 2I35) and activating conformation (PDB code: 3DQB)) were used for constructing GPHR homology models. The SYBYL program package was used for homology modeling and the AMBER 7.0 force field was applied for all energy and molecular dynamics calculations [10].
The 3D-structural morphings allow the visualization of a conformational transition between an inactivated and activated receptor conformation. User-driven analysis of specific amino acid changes, occurring during GPHR activation, help in clarifying wild type as well as

mutant functionalities. Kleinau et al. demonstrated that mapping of certain mutation phenotypes onto inactive vs. activated receptor states has been proven to be helpful in guiding and explaining experimental data in order to understand the molecular reasons for impaired receptor function [38].
Aside from 3D visualization, the focus was to develop an applet for analyzing the mutation dataset without knowledge of 3D-structural data interpretation, therefore a Snake-Plot-Tool and Designer are also provided. Both rely on the 2D representation of the GPHR sequence, highlight residues which have been subject to mutagenesis studies or are naturally occurring and hyperlink them to the published functional information. Customization of the snake-plots is possible by integrating own data and user-defined coloring.
In general, the linkage of functional data with structural information assists in acquiring deeper insight into the function and malfunction of disease-associated receptors. Thereby, we are able to contribute to new hypotheses regarding extracellular hormone binding, signal transduction to the intracellular coupled G protein and signaling regulation which might help to develop concepts for pharmacological interventions.

2.1.2 *Original publication: Research resource: Update and extension of a glycoprotein hormone receptors web application.*

Annika Kreuchwig, Gunnar Kleinau, Franziska Kreuchwig, Catherine L. Worth and Gerd Krause
Molecular Endocrinology April 2011, vol. 25, no. 4, 707-712
doi: 10.1210/me.2010-0510
The original article is available online at
`http://dx.doi.org/10.1210/me.2010-0510`

2.1.3 *Personal contribution*

For this study, I performed the major parts of the work including database development and dataset preparation, concept and design of the web application, implementation of tools and visualization options, with the exception of the 2D – Snake-Plot – Search of the web-application. This project was initiated during my Masters studies [41]. The major part of the database was finalized during the first year of my PhD studies. In addition, I wrote the manuscript together with Dr. Gunnar Kleinau and included suggestions from all co-authors.

Contributions of the authors:
Conception: G. Kleinau, A. Kreuchwig G. Krause
Design and development of database and web interface: A. Kreuchwig
Implementation of bioinformatic tools: A. Kreuchwig, F. Kreuchwig
Wrote the paper: A. Kreuchwig, G. Kleinau, G. Krause

2.2 Application: Analysis of naturally occurring mutations

2.2.1 *Synopsis*

One section of the SSFA-GPHR is dedicated to the collection and analysis of germline and somatic single side chain mutations. As described in the general introduction, GPCR side chain variants induce a variety of human diseases. Therefore, the deduction of the causal relationship between molecular dysfunction and patient phenotypes is extremely important. However, tools to analyze the molecular effects of genetic variations are thus far poorly provided. Therefore, a bioinformatic tool to interactively link structural and functional data has been developed. This helps to understand the molecular basis of naturally occurring gain- or loss-of-function mutations, is essential to guide the development of pharmacological treatments and paves the way to personalized medicine.
The interactive 3D Structural Search Tool projects naturally occurring mutations known for human TSHR, FSHR and LHCGR to the corresponding positions on their respective 3D structures. Positions where genetic variations are known to cause receptor defects are highlighted in accordance to their mutation phenotype (gain- or loss-of-function) in the Jmol structure viewer. The corresponding residues are linked with specific information such as the type of mutation, functional data (relative to the wild type) and the original source of publication. The information can be retrieved by clicking on the interactive links at the colored C-α atoms of the mutated positions. This allows investigation of spatial residue interactions and comparison with their respective functional data. Substitution of biophysically different side chains reveals insights in spatial and physical prerequisites at a certain position, for example the hydrophobic or hydrophilic character. Thereby they represent an ideal means in the identification of direct interaction partners either by directed functional interruption or maintenance of properties upon mutation.
Two examples of gain-of-function mutations are discussed in this publication. One relates to a pair of interacting amino acids (Val509 in TMH3 and Ala593 in TMH5), where mutagenesis experiments have shown that substitution of either side chain leads to an activating receptor phenotype. In the interactive 3D Structural Search Tool, the interplay of both residues can be analyzed in detail. This observation, together with experimental data, leads to the conclusion that this interaction is essential for wild type receptor function. The second example concentrates on interacting hydrophobic residues forming a hydrophobic patch to stabilize the basal TSHR conformation. Disruption of this motif of interacting hydrophobic amino acids (Leu512, Ile630 and Val597) located in helix 3, 5 and 6 by mutation causes constitutive TSHR receptor activity. The new tool is suitable in exploring and recognizing spatial interrelationships between amino acid side chains in combination with known variations which induce malfunctions in GPHRs. The mapping of functional data onto 3D structure can be very useful for researches from diverse fields. For example, clinicians may be interested in evaluating a patient's mutation to determine whether it is new or known and to access the published information in order to derive clues for possible treatment options. Moreover, researchers trying to reveal determinants for hormone binding and signal transduction are quite interested in identifying close interaction partners directly from the 3D structures, or

in evaluating the potential mechanisms responsible for the molecular defect caused by a specific mutation.

2.2.2 *Original publication: An interactive web-tool for molecular analyses links naturally occurring mutation data with three-dimensional structures of the rhodopsin-like glycoprotein hormone receptors.*

Gunnar Kleinau[#], **Annika Kreuchwig**[#], Catherine L. Worth and Gerd Krause
[#]These authors contributed equally to this work.
Human Mutation Jun 2010, 31, E1519-25
doi: 10.1002/humu.21265
The original article is available online at
http://dx.doi.org/10.1002/humu.21265

2.2.3 *Personal contribution*

This work is an application of the database for GPHRs focusing on a mutation dataset comprising of naturally occurring mutations in humans. Therefore, together with Dr. Gunnar Kleinau, we updated the dataset to include all recent publications of naturally occurring mutations in the SSFA-GPHR database. I incorporated the mutation data and for analysis of the latter, an interactive web-tool for linking structural and functional data was developed. Furthermore, my contributions also comprised of input in the discussion and conception of this study, and in writing approximately 50% of the manuscript. To show the usefulness of our new tool, several experimentally validated examples are discussed in the manuscript.

Contributions of the authors:
Conception: G. Kleinau, A. Kreuchwig, G. Krause
Performed analysis: G. Kleinau, A. Kreuchwig
Implementation of bioinformatic tools: A. Kreuchwig
Wrote the paper: G. Kleinau, A. Kreuchwig, CL. Worth, G. Krause

2.3 Application: activation induced spatial rearrangements of helices in GPHRs compared to other family A GPCRs

2.3.1 *Synopsis*

Signal transduction in GPCRs leads to a distinct spatial movement of specific transmembrane helices to one another. In particular, residues located in helices 5 and 6 are involved in stabilizing the basal or active state. The aim of this project was to investigate the interface between both helices in order to distinguish residues important for stabilizing the inactive state from residues enabling the activity switch. Therefore, results from site-directed mutagenesis, structural homology modeling and statistical sequence analysis of the interface residues were compared.

During template selection and multiple sequence alignment generation of family A GPCRs, we discerned that all published crystal structures so far, that are suitable as structural templates for homology modeling of GPHRs, contain a highly conserved proline at position 5.50 in transmembrane helix 5. However, the GPHRs contain an alanine at this position. Therefore, this deviation prompted to further analysis. Site-directed mutagenesis to proline, as well as to smaller and larger side chains followed by functional characterization, revealed the importance of this position in GPHRs. Side chain exchange to proline leads to a folding defect (strongly reduced cell surface expression, the receptor is probably unable to leave the early secretory pathway). Substitutions to glycine or valine lead to constitutive activation in the case of glycine, and decreased expression and basal activity for valine. Conclusions of these experiments can be drawn for the structural conformation of TMH5 of the TSHR and other GPHRs. It likely forms a different conformation as suggested by the crystal structures which have been published until date. They all demonstrate a widening of the helix induced by proline. Thus, for the TSHR and other GPCRs which have a substitution deviating from proline, a regular α-helix of TMH5 should be taken into account.

The next step was to characterize the residues lining the interface between helices 5 and 6. Therefore, the selection of amino acid positions for mutagenesis studies was guided by comparison with our database for GPHRs (`http://www.ssfa-gphr.de`). The functional relevance of ten so far uncharacterized positions was investigated by site-directed mutagenesis and functional characterization (cell surface expression, cAMP accumulation for basal and TSH-induced activity) to complete the data for the interface between TMH5 and TMH6. By linking these results with previous data collected in our SSFA-GPHR resource, as well as with structural information of the active and inactive receptor conformation, pairs of interacting residues either stabilizing the basal or active state can be defined.

Combination of the structural and functional results of this study revealed two important regions for TSHR activation. Firstly, a polar core region between TMH2, 3, 6 and 7 constraining the inactive state and secondly, hydrophobic interactions forming the interface between helices 5 and 6, stabilizing the active receptor state. Structural morphings help to visualize the dynamic process of helical rearrangements. In our web application for GPHRs, we now offer an applet to visually trace the helical shifts and side chain changes that potentially occur during TSHR activation. Additionally, a movie file which summarizes the results of

this study is provided as supplemental material.
To assess whether the conclusions drawn from this study are extendable to other family A GPCRs, we analyzed the sequence conservation of polar and non-polar residues in a set of 574 family A GPCRs. The frequencies of the two properties were calculated for each position of the transmembrane helices and subsequently plotted onto the crystal structure of bovine rhodopsin (PDB code: 1U19). In comparison to our TSHR-specific results, similar conclusions can be reached for family A GPCRs . Sequence analysis revealed: i.) a set of polar residues in TMHs 1, 2, 3 and 7 most likely involved in stabilizing the inactive state, and ii.) an interface between TMHs 3, 5 and 6 composed of non-polar residues which is presumably important for holding the active state. The conservation of these features throughout family A GPCRs suggests their fundamental role in regulating receptor activation.

2.3.2 *Original publication: From molecular details of the interplay between transmembrane helices of the thyrotropin receptor to general aspects of signal transduction in family A G-protein-coupled receptors (GPCRs).*

Gunnar Kleinau[#], Inna Hoyer[#], **Annika Kreuchwig**[#], Ann-Karin Haas, Claudia Rutz, Jens Furkert, Catherine L. Worth, Gerd Krause and Ralf Schülein
[#]These authors contributed equally to this work.
Journal of Biological Chemistry Jul 2011, 286, 25859-71.
doi: 10.1074/jbc.M110.196980
The original article is available online at
`http://dx.doi.org/10.1074/jbc.M110.196980`

2.3.3 *Personal contribution*

For this work, a simulation (morph) to illustrate the movements of all amino acids during the transition between basal and activated receptor conformation was created to support the experimental results of the study. Therefore, a Jmol applet to visualize and analyze the structural changes that occur during activation of the thyrotropin receptor was added. The database was used to design the study and analyze the results. Besides providing the new tool for our web application, a supplemental movie of the interpolation was generated using PyMOL Software and Python Scripts. Furthermore, my role included participation in the discussion and conception of this study, and writing the methods section used for simulation. Additionally, I wrote a description of the movie scenes for the supplemental material to highlight the importance of the new insights.

Contributions of the authors:
Conception: G. Kleinau, I. Hoyer, A. Kreuchwig, CL. Worth, G. Krause
Performed analysis: G. Kleinau, I. Hoyer, A. Kreuchwig, CL. Worth
Implementation of bioinformatic tools: A. Kreuchwig
Wrote the paper: G. Kleinau, I. Hoyer, A. Kreuchwig, CL. Worth, G. Krause

2.4 Web server for automatic template predictions and homology model building for G protein-coupled receptors

2.4.1 *Synopsis*

Homology model building of GPCRs highly relies on the template structure chosen. Available crystal structures represent only a small subset of all possible structural variations in family A GPCRs. They all possess a common general architecture, however, key sequence and structural differences can be identified in the published crystal structures. The five crystallized family A GPCR structures available until July 2010 (PDB codes: 1U19, 2Z73, 2VT4, 2RH1, 3EML) were analyzed for conserved and unique sequence motifs and structural features. Collectively, some GPCRs contain features that are present in different crystal structures. Using only one crystal structure as a single template, some structural features might be missed, suggesting that homology models should be built using a fragment-based approach whereby template selection is carried out for each transmembrane helix rather than the receptor as a whole. The presence or absence of specific structural features, such as the occurrence of prolines, disulphide bridges, insertions of amino acids and sequence similarity, supports the sequence-structure-based selection for the optimal template per helix. The selected helix fragments are then assembled by Modeller9v8 [20]. These assumptions were implemented in a workflow for identifying the appropriate template for each transmembrane helix [92]. The aim of this work was the automation of the workflow and application to a set of 5025 family A GPCR sequences, that were stored in the GPCRDB as of October 2009. The novelty and characteristic feature of GPCR-SSFE is that the automated workflow allows for the selection and assembly of several structural templates in a multi-template and fragment-based approach for modeling a given GPCR. This is particularly rather unique, in that it is the first method that utilizes the structural diversity of all available template structures by selecting the best matching template for each transmembrane helix instead of using one template for the entire receptor.

The GPCR-SSFE (`http://www.ssfa-7tmr.de/ssfe`) web application stores template predictions, sequence alignments, identified sequence and structure motifs, and homology models for 5025 family A GPCRs. Phylogenetic classification of the receptors stored in our database allows easy browsing of the dataset. Another way to retrieve the data is by searching for the UniProt entry name of a particular receptor. If the sequence of interest is not contained in the database, the user can submit a GPCR sequence for analysis and homology building to the web server. For template selection, the query sequence is aligned to a profile hidden markov model derived from the multiple sequence alignment of 54 family A GPCRs plus the five templates. The results include the final model which can be viewed using a Jmol applet and additionally providing information on the template-target alignments and template choice decisions. Furthermore, the results are available for download.

Following finalization of the web server and submission of the manuscript, two new family A GPCR crystal structure were published - the human dopamine D3 receptor (PDB code: 3PBL) and the CXCR4 chemokine receptor (PDB code: 3ODU). This provided a great opportunity to assess the quality of our fragment-based approach. For quality assessment, the backbone

RMSD (root-mean-square deviation) between our predicted and experimentally-determined structures, were calculated. Evaluation showed that the multiple template and fragment approach used in our web server is as accurate and in some cases superior to the models predicted by one of the current best performing homology modeling programs (I-TASSER, which was ranked first in CASP 10 (2012)). The resulting models offer a valuable starting point to evaluate mechanisms underlying structural or functional effects of mutations, or to guide structure-based drug design. The importance of the approach will further increase since the problem of template selection will become more complex when more crystal structures are solved.

2.4.2 *Original publication: GPCR-SSFE: a comprehensive database of G-protein-coupled receptor template predictions and homology models.*

Catherine L. Worth, **Annika Kreuchwig**, Gunnar Kleinau and Gerd Krause
BMC Bioinformatics May 2011, 23, 12:185.
doi: 10.1186/1471-2105-12-185
The original article is available online at
http://dx.doi.org/10.1186/1471-2105-12-185

2.4.3 *Personal contribution*

For consistency and recognizability, I designed a similar web interface according to the appearance of the SSFA-GPHR database. Additionally, I implemented the search function to retrieve results from the database holding all template predictions and homology models. Besides contributing to discussions and conception of the project, I constructed test cases for the workflow, analyzed the browser compatibility, and read and approved the manuscript.

Contributions of the authors:
Conception: CL. Worth, A. Kreuchwig, G. Kleinau, G. Krause
Algorithm development, homology modeling: CL. Worth
Implementation of a search function: A. Kreuchwig
Testing: A. Kreuchwig, G. Kleinau, G. Krause
Wrote the paper: CL. Worth

2.5 New structural information on the follitropin receptor offers new opportunities to describe thyrotropin receptor specific mechanisms

2.5.1 *Synopsis*

The N-terminal extracellular region (N-ECR) comprises the LRRD and hinge region. The LRRD has been solved fully or as fragments in hormone or autoantibody-bound state for the FSHR (FSHR LRRD in complex with FSH [PDB code: 1XWD]) and TSHR (TSHR LRRD in complex with either blocking [PDB code: 2XWT] or stimulating (PDB code: 3G04) autoantibody). Neither of these structures covered the full hormone binding site, including the hinge region, which has not been experimentally characterized by x-ray crystallography. Mutagenesis and binding studies have shown that it is involved in the regulation of signaling and plays an important role in hormone recognition and interaction. Until recently, no template structure was available to model these interactions for the TSHR between its hormone TSH and the hinge region, which connects the N-terminal LRRD with transmembrane helix 1. Jiang and co-workers published a complex structure comprising most parts of the N-ECR of FSHR and its hormone hFSH [32]. This complex now enables homology model building of the N-ECR, including, for the first time, the hinge region of the homologous TSHR/TSH complex.

The homology model of hTSHR is constructed by assembling fragments extracted from several template structures. For homology modeling of the N-terminal part of the leucine rich repeat domain, the nine repeats which were published earlier in an incomplete hTSHR antibody complex (PDB code: 3G04) were completed by a fragment of the recently published FSHR complex structure (PDB code: 4AY9) (TSHR positions Gln289-Ser304). Furthermore, the C-terminal hinge region (hTSHR positions Ser383-Ile411) is based on its corresponding region from this FSHR structure (positions Thr331-Ile359). Structural modifications and homology modeling procedures were performed with Sybyl X2.0, and energy minimization was performed by the AMBER F99 force field.

To validate the model, experimental data, including hormone binding (bTSH, hTSH, thyrostimulin), superagonistic effects, antibody interactions to specific residues in the hinge region, as well as signaling regulation, were evaluated. The model provides clues concerning residues involved in hormone binding and stabilizing the basal receptor conformation. In addition, a cluster of interacting residues at the interface between the hinge region and the hormone was identified, which is most likely involved in the conformational changes that occur at the N-ECR of GPHRs upon hormone binding. Most notably is a sulfated tyrosine (Tyr385) which appears to play a key role in hormone recognition and serves as a second hormone binding site. The negative Asp386 adjacent to the sulfated tyrosine 385 is involved in charge-charge interactions with the hormone. For antibody recognition another binding motif is proposed. In general, the autoantibody complex crystal structures reveal, that the interface is made up by the same general region of the LRRD, however as it is bulkier, it will not reach the second binding site around the sulfated Tyr385 at the hinge region, which appears to be glycoprotein hormone specific.

The systematic analyses of structural and functional data helped to provide new insights in TSHR-specific activation mechanisms. The complex model of TSHR/hTSH reveals distinct residues relevant for hormone recognition, which were previously partially characterized by mutagenesis studies to be in direct contact. Furthermore, patterns of interacting residues are shown to be involved in ligand-induced conformational changes, suggesting a possible movement of specific hinge region fragments, which eventually lead to receptor activation.

2.5.2 *Original publication: Extended and structurally supported insights into extracellular hormone binding, signal transduction and organization of the thyrotropin receptor.*

Gerd Krause, **Annika Kreuchwig** and Gunnar Kleinau
Plos One Dec 2012, 7, e52920.
doi: 10.1371/journal.pone.0052920
The original article is available online at
http://dx.doi.org/10.1371/journal.pone.0052920

2.5.3 *Personal contribution*

For this project I supported the homology model building process and assisted in the process of manuscript writing and image preparation.

Contributions of the authors:
Conception: G. Krause, A. Kreuchwig, G. Kleinau
Homology modeling: G. Krause, A. Kreuchwig, G. Kleinau
Analysis of TSHR-specific features: G. Krause, G. Kleinau
Wrote the paper: G. Krause, A. Kreuchwig, G. Kleinau

2.6 Novel structural insights bridge gaps in glycoprotein hormone receptor analyses - revision of the SSFA-GPHR and development of an interface analysis tool

2.6.1 *Synopsis*

The publication of the complex crystal structure of hFSHR with its bound hormone FSH required an update for the SSFA-GPHR web application (Section 2.1 on page 10). The homology modeling procedure described for the homologous hTSHR (see Section 2.5 on page 55) and drawn conclusions paved the way for an interactive interface analysis tool to enable the identification of contact points between N-ECR and hormone. These new insights together with the functional data stored in our web application, helps to identify essential residues for hormone recognition, binding as well as the molecular mechanisms for signal transduction.

The new crystal structure bridges a gap in the GPHRs, where a binding and signaling sensitive extracellular fragment, the hinge region (~100 residues) had been previously missing. Therefore, all previous published structures of the extracellular hormone binding site of GPHRs were not able to explain its role, because of the lack of structural knowledge concerning the second hormone binding site, located in the hinge region.

The novel structural findings demanded a revision of our web application SSFA-GPHR. To gain valuable knowledge about all receptor subtypes in complex with their hormones in our database, homology models were derived for TSHR and LHCGR, including different species (human and rat) for LHCGR and FSHR.

Additionally, a new interface analysis tool offers visualization and analysis options to illuminate the contacts and interactions between hormone and receptor. This not only allows comparison of binding patterns at GPHR subtypes, but also helps to decipher essential residues involved in ligand recognition or receptor activation. The user can visualize opposing binding patterns by selectively displaying the contact surface either for the hormone or receptor side. Moreover, distinct interaction partners or particular amino acids with specific biophysical properties can be depicted and highlighted for further analysis (e.g. positive vs. negative charges). For the contact analysis, surface rendering in the Jmol applet with distance mapping for contact analysis was combined. To distinguish the interactions, the color scale is chosen based on contact proximity of the van der Waals surfaces between potentially interacting partners such as from hydrogen bonds (red), to more distant ones (hydrophobic interactions, shown in yellow/green) (see Figure 2 in the original publication). This tool is extremely helpful in explaining the diverse GPHR functionalities, and in guiding comparative studies between the receptor subtypes to shed light on processes such as hormone binding and structural reorganization from the extracellular binding domain towards the intracellular G protein binding sites.

2.6.2 *Original publication: Research resource: novel structural insights bridge gaps in glycoprotein hormone receptor analyses.*

Annika Kreuchwig, Gunnar Kleinau and Gerd Krause
Molecular Endocrinology Aug 2013, 27(8):1357-63.
doi: 10.1210/me.2013-1115
The original article is available online at
http://dx.doi.org/10.1210/me.2013-1115

2.6.3 *Personal contribution*

For the update of the SSFA-GPHR resource, which was necessary due to the newly available structural information of the complex between the extracellular region of FSHR and its hormone FSH, I implemented all modifications in the database and web application. Moreover, I supported the homology building process. In accordance to the new structural information, the annotations in the alignments, numbering scheme and snake-plot designer were revised. The Jmol applet for visualizing functional information of the hormone-receptor complex were adapted and optimized. Furthermore, a new tool for the analysis and visualization of intermolecular interactions in the hormone-receptor complex was implemented. In addition, I contributed towards writing the manuscript and included suggestions from co-authors.

Contributions of the authors:
Conception: G. Krause, A. Kreuchwig, G. Kleinau
Homology modeling: G. Krause, A. Kreuchwig, G. Kleinau
Implementation: A.Kreuchwig
Wrote the paper: A. Kreuchwig

CHAPTER 3

Evolutionary approach: Linking sequence variation with structural information to reveal relevance of mutations

This chapter describes how the evolutionary relationship of GPCRs facilitates the evaluation of molecular effects caused by genetic variations. Firstly, it introduces the reader to the approach and its feasibility, which was verified by comparing the evolutionary conservation (77 ortholog sequences) with experimental data of a comprehensive in vitro mutant library (site-saturation mutagenesis of every possible substitution at 66 contiguous positions, in total 1254 mutants) [14]. Secondly, it covers the description and interpretation of the dataset which has been made available to the public as the $P2Y_{12}$ mutant library at `http://www.ssfa-7tmr.de/p2y12`.

3.1 Linking structural information and evolutionary sequence variation to reveal the functional relevance of GPCR variants

3.1.1 *Synopsis*

Evolutionary processes of conservation and natural selection leave a signature on protein sequences of extant species. Changes at individual positions raise the question of their functional relevance. Experimental testing of all possible gene variants at a particular position is either not, or only hardly feasible. For this reason, we analyzed whether the functional effect of mutations is predictable by evaluating a comparative ortholog sequence dataset including sequences (in total 77) from mammals, but also distantly related and evolutionary older species, such as fishes and amphibians. The rationale behind this approach is based on the fact that sequence diversity and conservation of a given protein is a result

of long evolutionary processes characterized by mutations and natural selection. Therefore, comparative sequence data should be able to identify distinct structural regions as well as provide insights into the spatial freedom and functional relevance of every amino acid position in a protein.

For analysis, a GPCR for ADP ($P2Y_{12}$ receptor) was chosen as an ideal candidate. Defects in this receptor cause severe impairment of platelet aggregation, frequently undergo natural selection, and are ultimately eliminated due to reduced viability. To evaluate the hypothesis whether the evolutionary conservation and variability left in the sequences of extant species (*in vivo*) provides insight into the functional relevance of every amino acid position a large ortholog dataset of 77 $P2Y_{12}$ species was compared to a set of *in vitro* functional data (site-saturation mutagenesis). Therefore, the functional effect of every possible substitution (19 variants) at each amino acid position within a portion of $P2Y_{12}$ (1254 mutants covering 66 contiguous receptor positions in the conserved transmembrane helices 6 and 7 as well as more variable ECL3) was determined. The mutants were expressed and functionally tested in a modified Saccharomyces cerevisiae system where receptor activity was directly coupled to yeast growth in a histidine-deficient medium. Activation of the mutant receptors was determined in the presence of the agonist 2-(methylthio)-ADP and further evaluated in 2-(methylthio)-ADP concentration-response (growth) curves.

For evaluation of the *in vivo* occurrence of variants based on the ortholog dataset and the *in vitro* function of the mutants a web-based system, the $P2Y_{12}$ mutant library (accessible at `http://www.ssfa-7tmr.de/p2y12`) was implemented. This database stores all experimental data and links the mutant information to data derived from sequences, sequence alignments and homology models of different activity states. Visualization and analysis features include interactive 2D- and 3D-structural analysis of mutations, as well as distinct search options for specific positions, spatial regions, or residue properties, and a quick alignment search. Additionally, a combined search providing information concerning the conservation or variation at multiple sites has been developed. This is a highly valuable tool, for example, in analyzing the conservation of the ligand binding site which usually involves co-evolving residues. The spatial distribution of these residues is visualized in a 3D interactive Jmol window and the conservation pattern is depicted in a multiple sequence alignment. The comparison of receptor function *in vivo* (multiple sequence alignment of extant species) and *in vitro* (functional characterization of mutants) in a pairwise manner, reveals a high correlation of 90.2% between both datasets. This high correlation confirms the hypothesis that ortholog sequence data are helpful to predict the functional relevance of individual positions and mutations for $P2Y_{12}$. The availability of a comprehensive ortholog sequence dataset may extend this approach to other GPCRs and conserved proteins as well.

The published work refers several times to supplementary material, in particular figures and tables, which are complementing the publication, see `http://www.fasebj.org/content/26/8/3273/suppl/DC1`.

3.1.2 *Original publication: Using ortholog sequence data to predict the functional relevance of mutations in G-protein-coupled receptors.*

Maxi Cöster, Doreen Wittkopf, **Annika Kreuchwig**, Gunnar Kleinau, Doreen Thor, Gerd Krause and Torsten Schöneberg
FASEB Journal Aug 2012, 26, 3273-81.
doi: 10.1096/fj.12-203737
The original article is available online at
http://dx.doi.org/10.1096/fj.12-203737

3.1.3 *Personal contribution*

For the collection, comparison and analysis of the *in vitro* functional characterization of all possible variants at 66 receptor positions at the $P2Y_{12}$ receptor, I implemented the database and web application. In this web-based system, we combine evolutionary sequence data, with experimental and structural information derived from homology models. For the web application, I implemented several tools for browsing, analyzing and comparing the experimental dataset by combining structural information and evolutionary sequence variation. The manuscript was primarily written in the laboratory of Prof. Dr. Torsten Schöneberg, but my contribution to the manuscript involved writing all the information concerning the web application, describing detailed possible applications and supporting the general process of manuscript preparation.

Contributions of the authors:
Conception: T. Schöneberg, M. Cöster, A. Kreuchwig, G. Kleinau, G.Krause
Homology modeling: G. Kleinau
Implementation: A.Kreuchwig
Statistical analysis: A.Kreuchwig, M.Cöster
Wrote the paper: M.Cöster, T.Schöneberg

CHAPTER 4

General Discussion, Future Perspectives and Conclusion

This chapter briefly recapitulates the achievements and results of this thesis. Furthermore, it will review how the outcome matches the objectives and the strategic approach set out in Chapter 1. An excursus will highlight collaborative projects from different research fields demonstrating how the methods presented in this thesis can be adapted for other proteins of interest. The chapter will also focus on future plans combining the concepts presented in this thesis to create an integrated platform for GPCR sequence-structure-function-analysis. Finally, the results will be evaluated to the recent developments in the field, providing a view on a broader context.

Great progress has been made over the past three decades in associating known disease phenotypes to their causing mutations in GPCRs. Dysregulation of GPCR function caused by variations, either leading to loss-of-function or gain-of-function phenotypes in patients or in *in vitro* studies, have contributed to our understanding of the pathophysiology of several GPCR-related diseases [83]. Recent crystallographic structures of different activity states, site-directed mutagenesis, structure-activity-relationships, evolutionary considerations, and other structural or functional studies of the GPCR, its ligand, G protein or β-arrestin interaction, start to clarify the plethora of structural and functional characteristics which are each individually susceptible to alterations caused by mutations. For the development of GPCR modulating drugs, it is essential to couple functional information derived from extensive mutagenesis studies which are unfortunately not uniformly published or deposited in a mutagenesis database with structural data from crystallographic, NMR spectroscopy or homology modeling techniques. However, while analyzing the structural data available thus far, one should consider that the static snapshots do not represent what is occurring within the cell, where these receptors are in constant flux. Therefore, these multifaceted approaches have to take the dynamic nature of GPCRs into account. Computational approaches, which

combine these multi-layered datasets, are necessary for yielding more efficient starting points for lead structure development by medicinal chemists.
The research described in this thesis provides novel web-based tools and databases which can be used interactively and are readily accessible to expert and non-expert users, facilitating visualization, analysis and extraction of data to guide future experimental studies and drug design. Linking the collection of mutation data with structural and evolutionary sequence information assists the progress for identifying functional residues important for ligand binding, receptor expression, G protein coupling, as well as β-arrestin interaction. The identification of tolerated modifications versus variations altering the receptor function together with ligand binding data reveal new and surprising roles of particular residues and might explain the varying responses of individuals to specific drugs. The work presented in this thesis concentrated on family A GPCRs, with a special focus on glycoprotein hormone receptors and P2Y nucleotide receptors. In particular, a web-based resource for the collection, analysis, comparison and visualization of mutation data in combination with sequence and structural information has been developed for the GPHRs (SSFA-GPHR, available at `http://www.ssfa-gphr.de`). During the course of the thesis, this database has been continuously developed by implementing novel tools, extending the dataset and visualization options to a valuable resource which is extensively used to guide or explain experimental studies. The database is constantly updated with novel structural insights, for example with the complex structure of the follitropin receptor and its hormone follitropin. This new structure permitted homology modeling for the homologous GPHRs, and the subsequent systematic analysis of structural and functional information from our resource revealed distinct residues crucial for TSH recognition and binding. These findings led to the implementation of a new interface visualization and analysis tool integrated in the SSFA-GPHR web application, allowing analysis of the contacts and interactions between hormones and their respective receptors, interactively.
A more general application was developed to cope with the template selection problem for GPCR homology modeling. Compared to the large number of receptors belonging to the family A GPCRs, the structural information available remains sparse. For this reason, homology models are the method of choice for generating 3D models of the GPCR of interest. To ease the template decision and allow the usage of multiple templates for modeling the structure of a given GPCR, a web-accessible pipeline, namely GPCR-Sequence-Structure-Feature-Extractor (SSFE, available at `http://www.ssfa-7tmr.de/ssfe`) for GPCR homology modeling, was developed. Besides querying the GPCR-SSFE for a certain GPCR, the web server with its associated database provides template predictions, sequence alignments, identified sequence and structure motifs and homology models for 5025 family A GPCRs. In comparison to other approaches, the novelty and uniqueness is that the template selection is carried out for each transmembrane helix rather than for the receptor as a whole and finally, the helix bundle is assembled in a multi-template and fragment based approach. This has the advantage of increased structural accuracy, due to the selection of the most appropriate template per helix of the entire set of available crystal structures, compared to the single template based approach. Thereby, specific sequence and structural features can be included to increase the feature and alignment coverage of the given GPCR. For benchmarking, newly

solved crystal structures of human dopamine D3 receptor (PDB code: 3PBL) and the CXCR4 chemokine receptor (PDB code: 3ODU) were compared to the homology models predicted by GPCR-SSFE, which did not contain these structures in the template pool. The results validated the effectiveness of the approach, yielding increased or at least similar accuracy in comparison with other available methods.
The vast amount of sequence data available due to the growing number of genome projects initiated new approaches to evaluate the functional importance of naturally occurring variants in proteins. A web-based system for the analysis of genotype-phenotype relationships by using both structural information and evolutionary sequence variation has been developed. For the initial proof of principle, an ADP-activated GPCR ($P2Y_{12}$) involved in platelet aggregation and target for the world's second highest selling drug, Clopidogrel, was selected [14]. The data generated for this study includes evolutionary information derived from 77 $P2Y_{12}$ sequences (orthologs), *in vitro* functional characterization of all 19 possible variants at 66 contiguous positions (1,254 mutants) and structural information through $P2Y_{12}$ homology models. These three components form the basis for the $P2Y_{12}$ mutant library (`http://www.ssfa-7tmr.de/p2y12/`). This web application provides several tools for browsing, analyzing and comparing the huge experimental dataset by combining structural information and evolutionary sequence variation. At the example of genetic variants of $P2Y_{12}$ known to cause platelet defects, we were able to highlight the utility of our approach. Therefore, we built initial homology models for all 77 $P2Y_{12}$ orthologs using our published workflow GPCR-SSFE. Comparison of sequence variation together with the space surrounding a specific variant in the structural alignment of all $P2Y_{12}$ homology models helped in evaluating the impact of particular mutational events. This might help to decipher the evolutionary constraints which determine the function of a protein, which will then facilitate the development of new strategies for pharmacological interventions, and in case of GPCRs, allow deeper insights to be gained into the mechanisms of ligand binding, signal transduction and regulation.

4.1 Excursus: collaborative projects predicting ligand-protein interactions supported by sequence-structure-function-analyses

During the time of this PhD thesis, collaborative projects were supported by sequence-structure-function analyses. Therefore, similar strategies as applied for GPCRs were used to support several studies for elucidating ligand-protein interactions. Here, two examples, both including great interdisciplinary work, are discussed. The first one, a collaborative effort between the Max Planck Institute for Infection Biology Berlin, Leibniz-Institut für Molekulare Pharmakologie (FMP) Berlin, Clinic for Pediatric Pneumology, Allergology and Neonatology, Hannover Medical School, Institute of Pharmacology and Structural Biology (IPBS), CNRS and University of Toulouse and German Rheumatism Research Centre Berlin (DRFZ), Berlin entitled "'AhR sensing of bacterial pigments orchestrates antibacterial defense"' has just been accepted by Nature [53]. In an essence, the aim of this project was to substantiate that the transcription factor, aryl hydrocarbon receptor (AhR), plays an important role in the immune defense system against bacteria. For proving this hypothesis

microbial insults, namely the phenazines from *Pseudomonas aeruginosa* and the naphthoquinone phthiocol from *Mycobacterium tuberculosis* were tested as ligands for this receptor, which has been shown to be sensible to structurally similar environmental pollutants, such as 2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD) [56]. To derive an interaction model of AhR with these bacterial virulence factors, a receptor model based on the C-terminal PAS domain of HIF-2 alpha [3] was modeled and subsequent *in silico* docking studies were performed. On the one hand, the AhR binding pocket is rather hydrophobic and contains possible proton donators: His291, Ser365 and Gln383. On the other hand the studied ligands (pyocyanin, 1-hydroxyphenazine, phenazine 1-carboxylic acid and phenazine 1-carboxamide from *P. aeruginosa* as well as the naphthoquinone phthiocol from *M. tuberculosis*), all share two proton acceptors and an aromatic ring system. The information about the ligands, functional data from AhR mutagenesis and TCDD binding studies in combination with structural information from the receptor model helped to identify the key residues for the *in silico* docking procedure and to suggest a mode of interaction. The *in silico* predictions were verified by a radioactive ligand binding assay.
The aim of the second collaboration was to derive competitive inhibitors of the interaction between A-kinase anchor proteins (AKAPs) and protein kinase A (PKA). Based on the NMR structures (PDB code: 1L6E and 2H9R) α-helical AKAP18 peptides were found to interact with the dimerization and docking (D/D) domain of PKA. As potential non-peptide mimetics for these α-helical peptides, *in silico* predictions led to the particular suggestion and subsequent synthesis of terpyridines. Docking of these terpyridines revealed a similar interaction profile as seen for AKAP18δ in the D/D domain of PKA characterized by predominately hydrophobic interactions, but also the formation of hydrogen bonds between the proton donators Gln4 and Gln14 in both helices of the D/D domain and the terpyridines docked in this study. The molecular modeling and docking studies were supported by NMR experiments validating the predictions [72].

4.2 Future Perspectives

4.2.1 *Extension of the web server for GPCR homology modeling*

During the last months extensive work has been carried out towards updating the GPCR-SSFE web server (implemented by Franziska Kreuchwig during the course of her Masters work) to include the newly published GPCR crystal structures, which have become available following the initial publication of the web server in 2011 [93]. The major aim of this update was the integration of representative crystal structures for all distinct rhodopsin-like GPCRs in the inactive state, which are available as of yet. On the one hand, novel ideas for database optimization and enhancement were developed, e.g. the implementation of a novel template selection strategy which is based on a fingerprint correlation scoring approach. On the other hand, the separate helix assembly for generating multi-template GPCR homology models in a fragment-based approach was retained, as recent studies have suggested that using multiple structures for homology model building instead of a single one leads to more accurate results for certain applications [5, 6, 46, 52]. Based on the positive experience and the results

achieved in the $P2Y_{12}$ project, another extension for the homology modeling web server has been planned and in part has commenced. For the analysis of sequence-structure-function relationships of family A GPCRs structures, homology models and functional data are important tools for understanding the receptor machinery. However, throughout evolution, GPCRs have evolved in many different ways finally providing us with a set of fine-tuned subtypes performing various functions in our body, each having their own set of ligands, activation mechanisms and transmitted signals. These evolutionary lines enable the comparison of species and GPCR subtypes. Therefore, our aim is to incorporate the evolutionary relationship of each GPCR by extending our dataset of 5025 GPCR to provide as many as 20 500 GPCR sequences, including many orthologs of each GPCR. This sequence dataset, together with the pre-calculation of homology models for the same number of GPCRs, will facilitate statistical analyses of variant and conserved positions at every helix position for a certain GPCR allowing the identification of functional residues for processes such as ligand binding, G protein- or β-arrestin interaction, which have remained unchanged throughout evolution. Besides further studying the evolutionary constraints which shaped GPCR function, it has become increasingly feasible to compare the structural differences of different activity states, due to the publication of antagonist-, partial agonist-, agonist-bound or even nanobody- and G protein-stabilized receptor complexes (reviewed in [85]). These active or active-intermediate structures cover only a small subset of family A GPCRs. To enable the comparison of different activity states and to allow the tracing of amino acid changes upon activation, an extension of the feature analysis to include published activated GPCR crystal structures has been planned. Additionally, tools for morphing in-between inactive and active receptor conformations are going to be implemented.

Furthermore, following pre-calculation of the homology models of 20 500 GPCRs, a binding pocket prediction and visualization tool would increase the informative value of our database and attract new users. Currently, different methods for pocket prediction are being tested, e.g. programs such as metaPocket-2 [98] or DoGSiteScorer [86] are taken into account for this analysis (metaPocket-2 combines eight pocket detection methods into one framework and defines consensus pockets to improve the prediction results). The algorithm with the best performance in terms of accuracy and reliability in finding the correct binding pocket will be successively implemented to visualize and compare different GPCRs binding pockets. This could be particularly useful in the identification of evolutionary constraints which can be derived from binding pocket comparisons of several, diverse GPCRs stored in the database. The accuracy of our web server compared to other existing methods has recently been evaluated by Latek and colleagues [46]. Only the omission of extra- and intracellular loops in the homology models provided by GPCR-SSFE has been criticized. Although thoroughly convinced that the loops are hard to predict and require preexisting knowledge or experimental constraints, a good solution was found for facilitating the loop modeling problem with a clean conscience. A cooperation with the group of Dr. Peter Hildebrand at the Charité Berlin led to the development of a modified version of their FragFit web server to permit the completion of the models by including the loops. This is achieved by an extension which will lead the user to a separate Jmol applet, where the optimal loop model for each loop (ECL1-3 and ICL1-3) may be chosen from a set of the ten highest scoring loop predictions

per loop. The user will interactively select the loops consecutively and finally decides for a final combination of loops best fitting to the model. With this extension, we will provide experienced users the opportunity to directly intervene in the homology modeling process. All ongoing projects or future ideas are summarized in Figure 4.1.
In April 2014, the COST Action CM1207 GLISTEN (GPCR-Ligand Interactions, Structures, and Transmembrane Signalling) which is a multidisciplinary European research network with the aim of joining forces in GPCR research, convened in Barcelona. During this meeting, the first GPCRDB satellite workshop led to a cooperative initiative to integrate existing datasets and web servers and develop novel ideas for GPCRDB tools. Therefore, the GPCRDB community should be expanded to bundle the strengths of various databases, web servers and programmers for further developments. The update and future plans of GPCR-SSFE generated great interest at the meeting and possibilities of integration into the GPCRDB were discussed. Furthermore, data population into the GPCRDB via reasonable interfaces of the GPHR (`http://www.ssfa-gphr.de`) and $P2Y_{12}$ (`http://www.ssfa-7tmr.de/p2y12`) datasets was a main issue at the workshop. The assembly of experts of a wide range of complementary GPCR research areas, from GPCRDB user to developer will guide the implementation and optimization of analysis and visualization tools aimed at providing a comprehensive platform for the whole GPCR community.

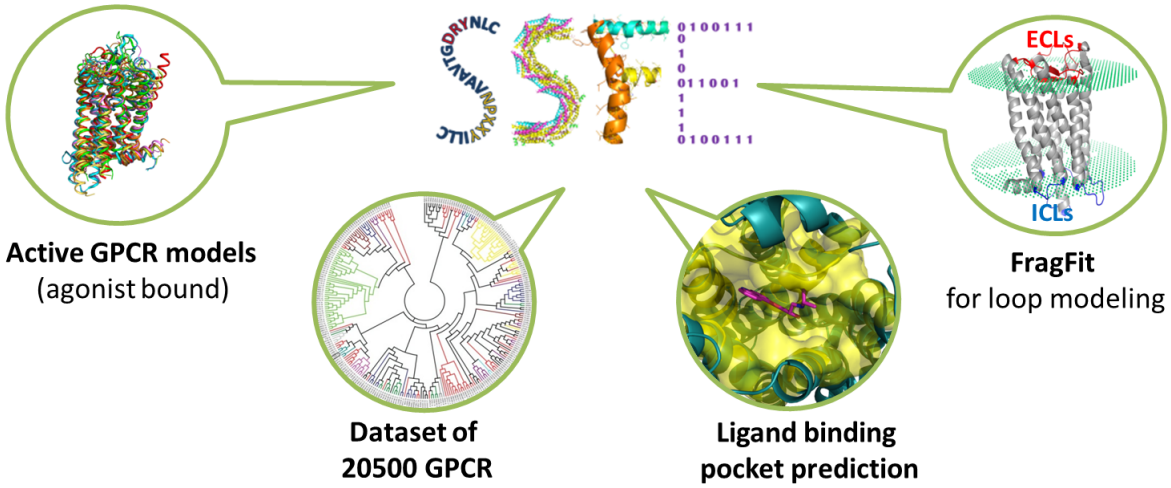

Figure 4.1 – Planned features for the SSFE web server.

Summary of future ideas for the GPCR-SSFE web server. In the next step, the homology modeling web server will be extended to create models of GPCRs in their active conformation. Furthermore, the dataset will be increased to 20 500 sequences including many orthologs. Another aim is to include binding site analysis, visualization and prediction tools. In addition, we aim to provide the user with the opportunity to build and choose their own set of loops for a particular model generated with our workflow.

4.2.2 *Integration of novel crystal structures for the ADP receptor*

Recently, the crystal structures of an antagonist-bound [PDB code: 4NTJ] and agonist-bound [PDB code: 4PXZ] conformation of the ADP receptor $P2Y_{12}$ have been published [94, 96].

The non-nucleotide antagonist AZD1283 holds the receptor in an inactive state with a distinct straight TMH5 conformation, which distinguishes $P2Y_{12}$ receptor from all other known class A GPCR structures. The nucleotide-agonist (2MeSADP) activates the receptor and reveals pronounced conformational changes between antagonist vs. agonist or non-nucleotide vs. nucleotide bound ligand complexes. Comparison of both structures demonstrates that the main structural rearrangements occur at the entry of the ligand binding pocket towards the extracellular region, which is in contrast to all other published active GPCR structures so far, which usually show large conformational changes in the intracellular arrangement of the helices to accommodate G proteins. Therefore, this new agonist-bound crystal structure may demonstrate an intermediate state where the agonist leads to an active conformation towards the binding pocket and extracellular region. For the intracellular region, an inactive conformation is quite likely, as stabilization by a G protein or G protein mimic is missing, and the helical positions observed resemble that of other intermediate crystal structures [70, 89]. However, these new crystal structures provide us with the opportunity to reassess our mutation data and to derive new insights for molecular reasons of how the ligand enters the receptor, how binding is coordinated, the reasoning behind mutations, and which changes occur upon activation inside the receptor, amongst others. The aim is to incorporate these novel structural information into our web-based $P2Y_{12}$ mutant library, in order to provide a resource to the community which offers numerous ways in extracting the available data. The availability of this resource assists in the guiding of decisions for drug discovery, personalized medicine or may facilitate the design of novel *in silico*, *in vitro* or *in vivo* experiments studying the underlying mechanisms of ligand binding, ligand effects, receptor activation and initiation of downstream effector cascades.

4.3 conclusion

The high degree of interdependency between structural, functional and evolutionary features clearly demands multi-layered systems biology approaches to efficiently handle the complexity associated with the current amount of vast data for family A GPCRs, and which will be provided in the future. The introduced web-based systems presented in this thesis attract users with differing background, including life science, bioinformatics and research-based pharmaceutical companies to study the molecular consequences of mutations in GPCRs. To guide these analyses, our web applications provide tools for i) identification of structurally and functionally important residues, ii) comparison of activated and inactivated receptor forms, iii) comparison of data in-between GPCR subtypes and even various species, iv) prediction of the functional relevance of either naturally or artificially occurring mutations, v) knowledge-driven multiple template based homology modeling and vi) interactive ligand-GPCR interface analysis. All these tools, which build on one another, are indispensable in deriving the causal link between genetic variation and altered receptor function which is essential for understanding GPCRs as a whole.

APPENDIX A

Appendix

A.1 Curriculum Vitae

In the following section I will provide my Curriculum Vitae, along with a publication list, as well as a list of my contributions to various conferences either as oral or poster presentation related to my PhD project.

The Curriculum Vitae is not included in the online version of this thesis due to privacy reasons.

Bibliography

[1] Ahmad, S. & Dray, A. (2004). Novel G protein-coupled receptors as pain targets. *Current opinion in investigational drugs (London, England : 2000)*, *5*(1), 67–70.

[2] Allan, C. M., Lim, P., Robson, M., Spaliviero, J., & Handelsman, D. J. (2009). Transgenic mutant D567G but not wild-type human FSH receptor overexpression provides FSH-independent and promiscuous glycoprotein hormone Sertoli cell signaling. *American journal of physiology. Endocrinology and metabolism*, *296*(5), E1022–8.

[3] Barouki, R., Aggerbeck, M., Aggerbeck, L., & Coumoul, X. (2012). The aryl hydrocarbon receptor system. *Drug metabolism and drug interactions*, *27*(1), 3–8.

[4] Benovic, J. L., Shorr, R. G., Caron, M. G., & Lefkowitz, R. J. (1984). The mammalian beta 2-adrenergic receptor: purification and characterization. *Biochemistry*, *23*(20), 4510–8.

[5] Bera, I., Laskar, A., & Ghoshal, N. (2011). Exploring the structure of opioid receptors with homology modeling based on single and multiple templates and subsequent docking: a comparative study. *Journal of molecular modeling*, *17*(5), 1207–21.

[6] Beuming, T. & Sherman, W. (2012). Current assessment of docking into GPCR crystal structures and homology models: successes, challenges, and guidelines. *Journal of chemical information and modeling*, *52*(12), 3263–77.

[7] Biebermann, H., Winkler, F., Handke, D., Teichmann, A., Gerling, B., Cameron, F., Eichhorst, J., Grüters, A., Wiesner, B., Kühnen, P., Krude, H., & Kleinau, G. (2012). New pathogenic thyrotropin receptor mutations decipher differentiated activity switching at a conserved helix 6 motif of family A GPCR. *The Journal of clinical endocrinology and metabolism*, *97*(2), E228–32.

[8] Boll, F. (1877). *Sull' anatomia e fisiologia della retina*. E. Loescher e Company.

[9] Caron, M. G. & Lefkowitz, R. J. (1976). Solubilization and characterization of the beta-adrenergic receptor binding sites of frog erythrocytes. *The Journal of biological chemistry*, *251*(8), 2374–84.

[10] Case, D. A. (2002). Molecular dynamics and NMR spin relaxation in proteins. *Accounts of chemical research*, *35*(6), 325–31.

[11] Cazzola, M., Page, C. P., Rogliani, P., & Matera, M. G. (2013). $\beta 2$-agonist therapy in lung disease. *American journal of respiratory and critical care medicine*, *187*(7), 690–6.

[12] Claus, M., Jaeschke, H., Kleinau, G., Neumann, S., Krause, G., & Paschke, R. (2005). A hydrophobic cluster in the center of the third extracellular loop is important for thyrotropin receptor signaling. *Endocrinology*, *146*(12), 5197–203.

[13] Conn, P. J., Christopoulos, A., & Lindsley, C. W. (2009). Allosteric modulators of GPCRs: a novel approach for the treatment of CNS disorders. *Nature reviews. Drug discovery*, *8*(1), 41–54.

[14] Cöster, M., Wittkopf, D., Kreuchwig, A., Kleinau, G., Thor, D., Krause, G., & Schöneberg, T. (2012). Using ortholog sequence data to predict the functional relevance of mutations in G-protein-coupled receptors. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, *26*(8), 3273–81.

[15] Dale, H. H. (1906). On some physiological actions of ergot. *The Journal of physiology*, *34*(3), 163–206.

[16] Davies, M. N., Secker, A., Halling-Brown, M., Moss, D. S., Freitas, A. A., Timmis, J., Clark, E., & Flower, D. R. (2008). GPCRTree: online hierarchical classification of GPCR function. *BMC research notes*, *1*(1), 67.

[17] De Lean, A., Stadel, J. M., & Lefkowitz, R. J. (1980). A ternary complex model explains the agonist-specific binding properties of the adenylate cyclase-coupled beta-adrenergic receptor. *The Journal of biological chemistry*, *255*(15), 7108–17.

[18] Douglas, S. D. & Leeman, S. E. (2011). Neurokinin-1 receptor: functional significance in the immune system in reference to selected infections and inflammation. *Annals of the New York Academy of Sciences*, *1217*, 83–95.

[19] Drake, M. T., Shenoy, S. K., & Lefkowitz, R. J. (2006). Trafficking of G protein-coupled receptors. *Circulation research*, *99*(6), 570–82.

[20] Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M.-Y., Pieper, U., & Sali, A. (2006). Comparative protein structure modeling using Modeller. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]*, *Chapter 5*, Unit 5.6.

[21] Filmore, D. (2004). It ' s a GPCR World. *Modern Drug Discovery*, (November), 24–27.

[22] Fredriksson, R., Lagerström, M. C., Lundin, L.-G., & Schiöth, H. B. (2003). The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Molecular pharmacology*, *63*(6), 1256–72.

[23] Freedman, N. J. & Lefkowitz, R. J. (1996). Desensitization of G protein-coupled receptors. *Recent progress in hormone research*, *51*, 319–51; discussion 352–3.

[24] Getting, S. J. (2006). Targeting melanocortin receptors as potential novel therapeutics. *Pharmacology & therapeutics*, *111*(1), 1–15.

[25] Haas, A.-K., Kleinau, G., Hoyer, I., Neumann, S., Furkert, J., Rutz, C., Schülein, R., Gershengorn, M. C., & Krause, G. (2011). Mutations that silence constitutive signaling activity in the allosteric ligand-binding site of the thyrotropin receptor. *Cellular and molecular life sciences : CMLS*, *68*(1), 159–67.

[26] Harmar, A. J., Hills, R. A., Rosser, E. M., Jones, M., Buneman, O. P., Dunbar, D. R., Greenhill, S. D., Hale, V. A., Sharman, J. L., Bonner, T. I., Catterall, W. A., Davenport, A. P., Delagrange, P., Dollery, C. T., Foord, S. M., Gutman, G. A., Laudet, V., Neubig, R. R., Ohlstein, E. H., Olsen, R. W., Peters, J., & Others (2009). IUPHAR-DB: the IUPHAR database of G protein-coupled receptors and ion channels. *Nucleic acids research*, *37*(Database issue), D680–5.

[27] Henry, H. L., Norman, A. W., Sierra, D. A., & Wilkie, T. M. (2003). Regulators of G-Protein Signaling (RGS) Superfamily. In *Encyclopedia of Hormones* (pp. 311–321). Elsevier Science Ltd.

[28] Herbert, J.-M. & Savi, P. (2003). P2Y12, a new platelet ADP receptor, target of clopidogrel. *Seminars in vascular medicine*, *3*(2), 113–22.

[29] Horn, F., Bettler, E., Oliveira, L., Campagne, F., Cohen, F. E., & Vriend, G. (2003). GPCRDB information system for G protein-coupled receptors. *Nucleic acids research*, *31*(1), 294–7.

[30] Horn, F., Weare, J., Beukers, M. W., Hörsch, S., Bairoch, A., Chen, W., Edvardsen, O., Campagne, F., & Vriend, G. (1998). GPCRDB: an information system for G protein-coupled receptors. *Nucleic acids research*, *26*(1), 275–9.

[31] Insel, P. A., Tang, C.-M., Hahntow, I., & Michel, M. C. (2007). Impact of GPCRs in clinical medicine: monogenic diseases, genetic variants and drug targets. *Biochimica et biophysica acta*, *1768*(4), 994–1005.

[32] Jiang, X., Liu, H., Chen, X., Chen, P.-H., Fischer, D., Sriraman, V., Yu, H. N., Arkinstall, S., & He, X. (2012). Structure of follicle-stimulating hormone in complex with the entire ectodomain of its receptor. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(31), 12491–6.

[33] Jmol 3D: an open-source Java viewer for chemical (2009). http://www.jmol.org/.

[34] Kazius, J., Wurdinger, K., van Iterson, M., Kok, J., Bäck, T., & Ijzerman, A. P. (2008). GPCR NaVa database: natural variants in human G protein-coupled receptors. *Human mutation*, *29*(1), 39–44.

[35] Khelashvili, G., Dorff, K., Shan, J., Camacho-Artacho, M., Skrabanek, L., Vroling, B., Bouvier, M., Devi, L. A., George, S. R., Javitch, J. A., Lohse, M. J., Milligan, G., Neubig, R. R., Palczewski, K., Parmentier, M., Pin, J.-P., Vriend, G., Campagne, F., & Filizola, M. (2010).

GPCR-OKB: the G Protein Coupled Receptor Oligomer Knowledge Base. *Bioinformatics (Oxford, England)*, *26*(14), 1804–5.

[36] Kleinau, G., Brehm, M., Wiedemann, U., Labudde, D., Leser, U., & Krause, G. (2007). Implications for molecular mechanisms of glycoprotein hormone receptors using a new sequence-structure-function analysis resource. *Molecular endocrinology (Baltimore, Md.)*, *21*(2), 574–80.

[37] Kleinau, G., Hoyer, I., Kreuchwig, A., Haas, A.-K., Rutz, C., Furkert, J., Worth, C. L., Krause, G., & Schülein, R. (2011). From molecular details of the interplay between transmembrane helices of the thyrotropin receptor to general aspects of signal transduction in family a G-protein-coupled receptors (GPCRs). *The Journal of biological chemistry*, *286*(29), 25859–71.

[38] Kleinau, G., Jaeschke, H., Mueller, S., Worth, C. L., Paschke, R., & Krause, G. (2008). Molecular and structural effects of inverse agonistic mutations on signaling of the thyrotropin receptor–a basally active GPCR. *Cellular and molecular life sciences : CMLS*, *65*(22), 3664–76.

[39] Kleinau, G., Kreuchwig, A., Worth, C. L., & Krause, G. (2010). An interactive web-tool for molecular analyses links naturally occurring mutation data with three-dimensional structures of the rhodopsin-like glycoprotein hormone receptors. *Human mutation*, *31*(6), E1519–25.

[40] Krause, G., Kreuchwig, A., & Kleinau, G. (2012). Extended and Structurally Supported Insights into Extracellular Hormone Binding, Signal Transduction and Organization of the Thyrotropin Receptor. *PLoS ONE*, *7*(12), e52920.

[41] Kreuchwig, A. (2009). *Development of specific Web Applications to link structural and functional data of Glycoprotein Hormone Receptors Master Thesis*. PhD thesis, Freie Universität Berlin.

[42] Kreuchwig, A., Kleinau, G., & Krause, G. (2013). Research resource: novel structural insights bridge gaps in glycoprotein hormone receptor analyses. *Molecular endocrinology (Baltimore, Md.)*, *27*(8), 1357–63.

[43] Kreuchwig, A., Kleinau, G., Kreuchwig, F., Worth, C. L., & Krause, G. (2011). Research resource: Update and extension of a glycoprotein hormone receptors web application. *Molecular endocrinology (Baltimore, Md.)*, *25*(4), 707–12.

[44] Langley, J. N. (1901). Observations on the physiological action of extracts of the suprarenal bodies. *The Journal of physiology*, *27*(3), 237–56.

[45] Lappano, R. & Maggiolini, M. (2011). G protein-coupled receptors: novel targets for drug discovery in cancer. *Nature reviews. Drug discovery*, *10*(1), 47–60.

[46] Latek, D., Pasznik, P., Carlomagno, T., & Filipek, S. (2013). Towards Improved Quality of GPCR Models by Usage of Multiple Templates and Profile-Profile Comparison. *PloS one*, *8*(2), e56742.

[47] Lefkowitz, R. J., Roth, J., & Pastan, I. (1970). Radioreceptor assay of adrenocorticotropic hormone: new approach to assay of polypeptide hormones in plasma. *Science (New York, N.Y.)*, *170*(3958), 633–5.

[48] Lefkowitz, R. J., Roth, J., Pricer, W., & Pastan, I. (1970). ACTH receptors in the adrenal: specific binding of ACTH-125I and its relation to adenyl cyclase. *Proceedings of the National Academy of Sciences of the United States of America*, *65*(3), 745–52.

[49] Liu, M., Zhao, J., Chen, K., Bian, X., Wang, C., Shi, Y., & Wang, J. M. (2012). G protein-coupled receptor FPR1 as a pharmacologic target in inflammation and human glioblastoma. *International immunopharmacology*, *14*(3), 283–8.

[50] Maeda, T., Imanishi, Y., & Palczewski, K. (2003). Rhodopsin phosphorylation: 30 years later. *Progress in retinal and eye research*, *22*(4), 417–34.

[51] Martí-Solano, M., Guixà-González, R., Sanz, F., Pastor, M., & Selent, J. (2013). Novel insights into biased agonism at G protein-coupled receptors and their potential for drug design. *Current pharmaceutical design*, *19*(28), 5156–66.

[52] Mobarec, J. C., Sanchez, R., & Filizola, M. (2009). Modern homology modeling of G-protein coupled receptors: which structural template to use? *Journal of medicinal chemistry*, *52*(16), 5207–16.

[53] Moura-Alves, P., Faé, K., Houthuys, E., Dorhoi, A., Kreuchwig, A., Furkert, J., Barison, N., Diehl, A., Munder, A., Constant, P., Skrahina, T., Guhlich-Bornhof, U., Klemm, M., Koehler, A.-B., Bandermann, S., Goosmann, C., Mollenkopf, H.-J., Hurwitz, R., & Others (2014). AhR sensing of bacterial pigments orchestrates antibacterial defense. *Nature*, *accepted*.

[54] Nathans, J. & Hogness, D. S. (1983). Isolation, sequence analysis, and intron-exon arrangement of the gene encoding bovine rhodopsin. *Cell*, *34*(3), 807–14.

[55] Nguyen, C. A., Akiba, Y., & Kaunitz, J. D. (2012). Recent advances in gut nutrient chemosensing. *Current medicinal chemistry*, *19*(1), 28–34.

[56] Nguyen, L. P. & Bradfield, C. a. (2008). The search for endogenous activators of the aryl hydrocarbon receptor. *Chemical research in toxicology*, *21*(1), 102–16.

[57] Nishihara, E., Chen, C.-R., Mizutori-Sasai, Y., Ito, M., Kubota, S., Amino, N., Miyauchi, A., & Rapoport, B. (2012). Deletion of thyrotropin receptor residue Asp403 in a hyperfunctioning thyroid nodule provides insight into the role of the ectodomain in ligand-induced receptor activation. *Journal of endocrinological investigation*, *35*(1), 49–53.

[58] Nurden, P., Savi, P., Heilmann, E., Bihour, C., Herbert, J. M., Maffrand, J. P., & Nurden, A. (1995). An inherited bleeding disorder linked to a defective interaction between ADP and its receptor on platelets. Its influence on glycoprotein IIb-IIIa complex function. *The Journal of clinical investigation*, *95*(4), 1612–22.

[59] Overington, J. P., Al-Lazikani, B., & Hopkins, A. L. (2006). How many drug targets are there? *Nature reviews. Drug discovery*, *5*(12), 993–6.

[60] Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M., & Miyano, M. (2000). Crystal structure of rhodopsin: A G protein-coupled receptor. *Science (New York, N.Y.)*, *289*(5480), 739–45.

[61] Palczewski, K., McDowell, J. H., Jakes, S., Ingebritsen, T. S., & Hargrave, P. A. (1989). Regulation of rhodopsin dephosphorylation by arrestin. *The Journal of biological chemistry*, *264*(27), 15770–3.

[62] Pan, H.-L., Wu, Z.-Z., Zhou, H.-Y., Chen, S.-R., Zhang, H.-M., & Li, D.-P. (2008). Modulation of pain transmission by G-protein-coupled receptors. *Pharmacology & therapeutics*, *117*(1), 141–61.

[63] Park, J. H., Scheerer, P., Hofmann, K. P., Choe, H.-W., & Ernst, O. P. (2008). Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature*, *454*(7201), 183–7.

[64] Rasmussen, S. r. G. F., DeVree, B. T., Zou, Y., Kruse, A. C., Chung, K. Y., Kobilka, T. S., Thian, F. S., Chae, P. S., Pardon, E., Calinski, D., Mathiesen, J. M., Shah, S. T. A., Lyons, J. A., Caffrey, M., Gellman, S. H., Steyaert, J., Skiniotis, G., Weis, W. I., Sunahara, R. K., & Kobilka, B. K. (2011a). Crystal structure of the $\beta 2$ adrenergic receptor-Gs protein complex. *Nature*, *477*(7366), 549–55.

[65] Rasmussen, S. r. G. F., DeVree, B. T., Zou, Y., Kruse, A. C., Chung, K. Y., Kobilka, T. S., Thian, F. S., Chae, P. S., Pardon, E., Calinski, D., Mathiesen, J. M., Shah, S. T. A., Lyons, J. A., Caffrey, M., Gellman, S. H., Steyaert, J., Skiniotis, G., Weis, W. I., Sunahara, R. K., & Kobilka, B. K. (2011b). Crystal structure of the $\beta 2$ adrenergic receptor-Gs protein complex. *Nature*, *477*(7366), 549–55.

[66] Rivero-Müller, A., Chou, Y.-Y., Ji, I., Lajic, S., Hanyaloglu, A. C., Jonas, K., Rahman, N., Ji, T. H., & Huhtaniemi, I. (2010). Rescue of defective G protein-coupled receptor function in vivo by intermolecular cooperation. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(5), 2319–24.

[67] Robison, G. A., Butcher, R. W., & Sutherland, E. W. (1967). Adenyl cyclase as an adrenergic receptor. *Annals of the New York Academy of Sciences*, *139*(3), 703–23.

[68] Römpler, H., Stäubert, C., Thor, D., Schulz, A., Hofreiter, M., & Schöneberg, T. (2007). G protein-coupled time travel: evolutionary aspects of GPCR research. *Molecular interventions*, *7*(1), 17–25.

[69] Rosenbaum, D. M., Cherezov, V., Hanson, M. A., Rasmussen, S. r. G. F., Thian, F. S., Kobilka, T. S., Choi, H.-J., Yao, X.-J., Weis, W. I., Stevens, R. C., & Kobilka, B. K. (2007). GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function. *Science (New York, N.Y.)*, *318*(5854), 1266–73.

[70] Rosenbaum, D. M., Zhang, C., Lyons, J. A., Holl, R., Aragao, D., Arlow, D. H., Rasmussen, S. r. G. F., Choi, H.-J., Devree, B. T., Sunahara, R. K., Chae, P. S., Gellman, S. H., Dror, R. O., Shaw, D. E., Weis, W. I., Caffrey, M., Gmeiner, P., & Kobilka, B. K. (2011). Structure and function of an irreversible agonist-β(2) adrenoceptor complex. *Nature*, *469*(7329), 236–40.

[71] Satagopam, V. P., Theodoropoulou, M. C., Stampolakis, C. K., Pavlopoulos, G. A., Papandreou, N. C., Bagos, P. G., Schneider, R., & Hamodrakas, S. J. (2010). GPCRs, G-proteins, effectors and their interactions: human-gpDB, a database employing visualization tools and data integration techniques. *Database : the journal of biological databases and curation*, *2010*, baq019.

[72] Schäfer, G., Milić, J., Eldahshan, A., Götz, F., Zühlke, K., Schillinger, C., Kreuchwig, A., Elkins, J. M., Abdul Azeez, K. R., Oder, A., Moutty, M. C., Masada, N., Beerbaum, M., Schlegel, B., Niquet, S., Schmieder, P., Krause, G., & Others (2013). Highly functionalized terpyridines as competitive inhibitors of AKAP-PKA interactions. *Angewandte Chemie (International ed. in English)*, *52*(46), 12187–91.

[73] Scheerer, P., Park, J. H., Hildebrand, P. W., Kim, Y. J., Krauss, N., Choe, H.-W., Hofmann, K. P., & Ernst, O. P. (2008). Crystal structure of opsin in its G-protein-interacting conformation. *Nature*, *455*(7212), 497–502.

[74] Schertler, G. F., Villa, C., & Henderson, R. (1993). Projection structure of rhodopsin. *Nature*, *362*(6422), 770–2.

[75] Schöneberg, T., Schulz, A., Biebermann, H., Hermsdorf, T., Römpler, H., & Sangkuhl, K. (2004). Mutant G-protein-coupled receptors as a cause of human diseases. *Pharmacology & therapeutics*, *104*(3), 173–206.

[76] Sharman, J. L., Benson, H. E., Pawson, A. J., Lukito, V., Mpamhanga, C. P., Bombail, V., Davenport, A. P., Peters, J. A., Spedding, M., & Harmar, A. J. (2013). IUPHAR-DB: updated database content and new features. *Nucleic acids research*, *41*(Database issue), D1083–8.

[77] Small, K. M., Tanguay, D. A., Nandabalan, K., Zhan, P., Stephens, J. C., & Liggett, S. B. (2003). Gene and protein domain-specific patterns of genetic variability within the G-protein coupled receptor superfamily. *American journal of pharmacogenomics : genomics-related research in drug development and clinical practice*, *3*(1), 65–71.

[78] Strotmann, R., Schröck, K., Böselt, I., Stäubert, C., Russ, A., & Schöneberg, T. (2011). Evolution of GPCR: change and continuity. *Molecular and cellular endocrinology*, *331*(2), 170–8.

[79] Tan, Q., Zhu, Y., Li, J., Chen, Z., Han, G. W., Kufareva, I., Li, T., Ma, L., Fenalti, G., Li, J., Zhang, W., Xie, X., Yang, H., Jiang, H., Cherezov, V., Liu, H., Stevens, R. C., Zhao, Q., & Wu, B. (2013). Structure of the CCR5 Chemokine Receptor-HIV Entry Inhibitor Maraviroc Complex. *Science (New York, N.Y.)*, *341*(6152), 1387–1390.

[80] Tao, Y.-X. (2006). Inactivating mutations of G protein-coupled receptors and diseases: structure-function insights and therapeutic implications. *Pharmacology & therapeutics*, *111*(3), 949–73.

[81] "The Nobel Prize in Chemistry 2012 - Popular Information". Nobelprize.org (2012). Cells and sensibility.

[82] Van Durme, J., Horn, F., Costagliola, S., Vriend, G., & Vassart, G. (2006). GRIS: glycoprotein-hormone receptor information system. *Molecular endocrinology (Baltimore, Md.)*, *20*(9), 2247–55.

[83] Vassart, G. & Costagliola, S. (2011). G protein-coupled receptors: mutations and endocrine diseases. *Nature reviews. Endocrinology*, *7*(6), 362–72.

[84] Vassilatis, D. K., Hohmann, J. G., Zeng, H., Li, F., Ranchalis, J. E., Mortrud, M. T., Brown, A., Rodriguez, S. S., Weller, J. R., Wright, A. C., Bergmann, J. E., & Gaitanaris, G. A. (2003). The G protein-coupled receptor repertoires of human and mouse. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(8), 4903–8.

[85] Venkatakrishnan, a. J., Deupi, X., Lebon, G., Tate, C. G., Schertler, G. F., & Babu, M. M. (2013). Molecular signatures of G-protein-coupled receptors. *Nature*, *494*(7436), 185–194.

[86] Volkamer, A., Kuhn, D., Rippmann, F., & Rarey, M. (2012). DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics (Oxford, England)*, *28*(15), 2074–5.

[87] Vroling, B., Sanders, M., Baakman, C., Borrmann, A., Verhoeven, S., Klomp, J., Oliveira, L., de Vlieg, J., & Vriend, G. (2011). GPCRDB: information system for G protein-coupled receptors. *Nucleic acids research*, *39*(Database issue), D309–19.

[88] Wald, G. (1933). Vitamin A in the retina. *Nature*, *132*, 316–317.

[89] Warne, T., Moukhametzianov, R., Baker, J. G., Nehmé, R., Edwards, P. C., Leslie, A. G. W., Schertler, G. F. X., & Tate, C. G. (2011). The structural basis for agonist and partial agonist action on a β(1)-adrenergic receptor. *Nature*, *469*(7329), 241–4.

[90] Winzell, M. S. & Ahrén, B. (2007). G-protein-coupled receptors and islet function-implications for treatment of type 2 diabetes. *Pharmacology & therapeutics*, *116*(3), 437–48.

[91] Wolfe, B. L. & Trejo, J. (2007). Clathrin-dependent mechanisms of G protein-coupled receptor endocytosis. *Traffic (Copenhagen, Denmark)*, *8*(5), 462–70.

[92] Worth, C. L., Kleinau, G., & Krause, G. (2009). Comparative sequence and structural analyses of G-protein-coupled receptor crystal structures and implications for molecular models. *PloS one*, *4*(9), e7011.

[93] Worth, C. L., Kreuchwig, A., Kleinau, G., & Krause, G. (2011). GPCR-SSFE: a comprehensive database of G-protein-coupled receptor template predictions and homology models. *BMC bioinformatics*, *12*, 185.

[94] Zhang, J., Zhang, K., Gao, Z.-G., Paoletta, S., Zhang, D., Han, G. W., Li, T., Ma, L., Zhang, W., Müller, C. E., Yang, H., Jiang, H., Cherezov, V., Katritch, V., Jacobson, K. a., Stevens, R. C., Wu, B., & Zhao, Q. (2014). Agonist-bound structure of the human P2Y12 receptor. *Nature*, *509*(7498), 119–122.

[95] Zhang, J. & Zhang, Y. (2010). GPCRRD: G protein-coupled receptor spatial restraint database for 3D structure modeling and function annotation. *Bioinformatics (Oxford, England)*, *26*(23), 3004–5.

[96] Zhang, K., Zhang, J., Gao, Z.-G., Zhang, D., Zhu, L., Han, G. W., Moss, S. M., Paoletta, S., Kiselev, E., Lu, W., Fenalti, G., Zhang, W., Müller, C. E., Yang, H., Jiang, H., Cherezov, V., Katritch, V., Jacobson, K. A., Stevens, R. C., Wu, B., & Zhao, Q. (2014). Structure of the human P2Y12 receptor in complex with an antithrombotic drug. *Nature*, *509*(7498), 115–118.

[97] Zhang, Y., Devries, M. E., & Skolnick, J. (2006). Structure modeling of all identified G protein-coupled receptors in the human genome. *PLoS computational biology*, *2*(2), e13.

[98] Zhang, Z., Li, Y., Lin, B., Schroeder, M., & Huang, B. (2011). Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. *Bioinformatics (Oxford, England)*, *27*(15), 2083–8.

Declaration

Herewith I affirm that I have written this thesis entitled *"Functional relevance of genetic variants in GPCRs"* on my own. Cited sources of literature are perceptibly marked and listed at the end of this thesis.

The work was not submitted previously in same or similar form to another examination committee and was not yet published, unless stated.

Annika Kreuchwig, Berlin, July 2014